

Wed May 12 09:50:20 2004

us-10-043-774b-1.rng

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 09:21:06 ; Search time 921 seconds  
(without alignments)

10738.116 Million cell updates/sec

Title: US-10-043-774B-1

Sequence: 1 atgatacttgcgtgacac.....acgccttactactctctaa 2328

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2328	100.0	2328	9	ADB94986 Human hae
2	2311	99.3	3472	3	AAAO7588 Human plw
3	2200.2	94.5	2579	3	ADB63315 Human CDN
4	1680.6	72.2	4064	3	AAAO7587 Mouse plw
5	1569	67.4	2292	6	ABV89287 Human COL
6	680	29.2	3138	9	ADB63522 Human CDN
7	458.4	19.7	2009	6	ABO61109 Human mtn
8	458.4	19.7	2009	6	ACA62542 Human ful
9	435.2	18.7	3620	7	ADA53334 Human cod
10	395.4	17.0	423	3	AAFI1880 Human bre
11	381	16.4	3649	5	AAAI3630 CDNA sequ
12	365.4	15.7	367	7	ABZ19511 Group III
13	364	15.6	367	7	ABZ20005 Group III
14	330	14.2	2272	4	AAH15960 Human CDN
15	310.6	13.3	2838	3	ABLI10571 Drosophila
16	286.4	12.3	300	2	AAZ13404 Human gen
17	263.4	11.3	3524	4	ABLI10535 Drosophila
18	254.6	11.0	3047	3	AAAO7586 Drosophila
19	254.6	10.9	1256	9	AADE3945 Human pro
20	237.2	10.2	1733	4	AAH14032 Human CDN
21	207.4	8.9	209	7	ABZ19102 Group III
22	184	7.9	5671	4	ABLI10570 Drosophila
23	171.8	7.4	548	6	ABN62610 Human can

24	149.2	6.4	1026	5	AAST70144 DNA encod
25	138	5.9	7667	4	ABLI10534 Drosophila
26	113.2	4.9	786	4	AAH05561 Human CDN
27	111.2	4.8	2914	4	AAK53425 Human pol
28	111.2	4.8	2914	7	ABX13625 Human cyt
29	111.2	4.8	3050	4	AAH16058 Human CDN
30	111.2	4.8	3325	7	AAH05860 Human nuc
31	111.2	4.8	3756	7	ABX34752 Human mtd
32	108.8	4.7	570	8	ACA62541 Human par
33	104.2	4.5	325	4	AAI84876 Human par
34	95.6	4.1	444	4	AAST7327 CDNA #3 e
35	91.2	3.9	601	5	AAST70143 DNA encod
36	90	3.9	646	5	AAST7101 DNA encod
37	88.2	3.8	7478	9	ADB81486 Human CDN
38	86.6	3.7	3996	4	AAH14510 Human CDN
39	84	3.6	1238	8	ADB17487 Rice post
40	83.2	3.6	2951	4	AAH24562 Translati
41	83	3.5	490	8	ACH23411 Human adu
42	82	3.5	642	4	AAH07789 Human CDN
43	82	3.5	1968	4	AAH14966 Human CDN
44	81.8	3.5	3011	6	ABQ93304 Human cod
45	81.8	3.5	3580	6	ABST8717 Human CDN

## ALIGNMENTS

RESULT 1  
ADB94986 standard; CDNA; 2328 BP.  
ID ADB94986;  
AC ADB94986;  
DT 04-DEC-2003 (first entry)  
XX Human haematopoietic growth regulatory protein, H1W1, CDNA.  
XX Human; sv; gene; haematopoietic growth regulatory protein; H1W1;  
XX Cytosolic; stem cell therapy; bone marrow; peripheral blood;  
XX CD4+ haematopoietic stem cell; leukaemia; cancer.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS 1..2328 /\*tag= a  
FT /product= "H1W1"  
XX US2003129599-A1.  
XX 10-JUL-2003.  
XX 10-JAN-2002; 2002US-00043774.  
XX 10-JAN-2002; 2002US-00043774.  
XX (SHAR/) SHARMA A.  
XX (HOFF/) HOFFMAN R.  
XX Sharma A, Hoffman R;  
XX WPI; 2003-744346/70.  
XX P-PSDB; ADB94987.  
XX New h1w1 nucleic acids, for screening compounds which modulate h1w1,  
XX useful as cancer therapeutic.  
XX Claim 4; Fig 1; 39pp; English.  
XX The invention relates to a nucleic acid comprising a nucleotide sequence  
XX encoding a human h1w1 protein (a haematopoietic growth regulatory gene).  
XX Also included are the h1w1 protein, the h1w1 CDNA (used as a probe), a  
XX recombinant expression construct comprising the h1w1 CDNA, a cell culture



[illegible]

XX	Claim 19; Page 189-194; 201pp; English.
PS	This sequence encodes the human p1w1 family protein, designated h1w1. The
CC	p1w1 family nucleic acids and polypeptides are used in gene therapy of
CC	diseases such as cancer and also in various research and diagnostic
CC	applications. The sequences can also be used to treat tissue dystrophy,
CC	anaemia, immunodeficiency, and male infertility
XX	
SQ	Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 0 U; 3 Other;
Query Match	99.3%; Score 2311; DB 3; Length 3472;
Best Local Similarity	99.7%; Pred. No. 0;
Matches 2314; Conservative	0; Mismatches 7; Indels 0; Gaps 0;
QY	8 TTGCTGTATACAGCAGCGAAGCTTAAAGCATGTTAAAGAAATCAAAAACAGTTCCTTAG
DB	432 TTGGTGTAATAACAAGGACGAAGCTTAAGCATGTTAAAGAAATCAAAAACAGTTCCTTAG
QY	68 GCATTATAGTAAGTTAAGCACTAACCATTTCCGGCTGCACATCCCGGCCAGTGGGCT
DB	492 GCATTATAGTAAGTTAAGCACTAACCATTTCCGGCTGCACATCCCGGCCAGTGGGCT
QY	128 TATATCAGATACCATCTACTATAACCACATGATGSAAGCCAGAAGACTCCGTTCACTC
DB	552 TATATCAGATACCATCTACTATAACCACATGATGSAAGCCAGAAGACTCCGTTCACTC
QY	188 TTCCTTTTTCAAACGAGAGATCTAATGGAAAGTGTCACTGCTTTGATGGAACGATATAT
DB	612 TTCCTTTTTCAAACGAGAGATCTAATGGAAAGTGTCACTGCTTTGATGGAACGATATAT
QY	248 TTTTACCTTAAAGACTACAGCAAAAAGTTACTGAAGTTTATAGTAAGACCGGATGAG
DB	672 TTTTACCTTAAAGACTACAGCAAAAAGTTACTGAAGTTTATAGTAAGACCGGATGAG
QY	308 AGAGATGAGAGTAAAGATCACTTTACAAATGAACTTCCAACCTACATCACCACTTGT
DB	732 AGAGATGAGAGTAAAGATCACTTTACAAATGAACTTCCAACCTACATCACCACTTGT
QY	368 TGCAATGCTATATATATTTTACAGGAGCTTTGAAATCATGATTTGACAACAATTG
DB	792 TGCAATGCTATATATATTTTACAGGAGCTTTGAAATCATGATTTGACAACAATTG
QY	428 GACGAAATATATTAACCCAAATGACCCAAATGATTTCCAAAGTCAACGCTTGATTT
DB	852 GACGAAATATATTAACCCAAATGACCCAAATGATTTCCAAAGTCAACGCTTGATTT
QY	488 GGCTCGGCTCACTACTTCCATCCCTTCAGTATGAAAGAGATCATGCTGTCAGTGA
DB	912 GGCTCGGCTCACTACTTCCATCCCTTCAGTATGAAAGAGATCATGCTGTCAGTGA
QY	548 TTAGCATTAAGTCCCTGGAAGTGAAGCTGTTTGGATTTCAAGTTCACCTTTATCATC
DB	972 TTAGCATTAAGTCCCTGGAAGTGAAGCTGTTTGGATTTCAAGTTCACCTTTATCATC
QY	608 AGACGAGAGACATTAATTTTCAAGAACAAAGTTCCAAAGACTATAGGTTAGTGTTC
DB	1032 AGACGAGAGACATTAATTTTCAAGAACAAAGTTCCAAAGAAATTAAGTITTAGTGTTC
QY	668 TTACCAAGTATACATTAAGACATACAGAGTGAATGATATTGACTGGACCAAGATCCCA
DB	1092 TTACCAAGTATACATTAAGACATACAGAGTGAATGATATTGACTGGACCAAGATCCCA
QY	728 AAGACACTTTTAAAGAACCCGAGCTCTGAAGTCAAGCTTCTTAATATCTACAGAAAGC
DB	1152 AAGACACTTTTAAAGAACCCGAGCTCTGGGGTCAAGCTTCTTAATATCTACAGAAAGC
QY	788 AATACAAACCAAGATCAACGACTTGAAGCAAGCTGCTTGGTCAACGACCCCAAGANA
DB	1212 AATACAAACCAAGATCAACGACTTGAAGCAAGCTGCTTGGTCAACGACCCCAAGANA
QY	848 GAGCGAGGCCCTGGGGGACACTGCGAAGGCTCGCAATGCTCATCTGAGCTCTGTATC

Db 1272 GGGGGGGCCCTGGGGGGGACACTGCGAGGCGCTGCATGCTCATTTCTGAGCTCTGCTATC 1331  
 QY 908 TTACAGGCTTACCTGATTAATAATGGGTAAATATTTTAACTGATGAAGAACTTAGACGCTTC 967  
 Db 1332 TTACAGGCTTACCTGATTAATAATGGGTAAATATTTTAACTGATGAAGAACTTAGACGCTTC 1391  
 QY 968 ATACAGGCTTACCTGATTAATAATGGGTAAATATTTTAACTGATGAAGAACTTAGACGCTTC 1027  
 Db 1392 ATACAGGCTTACCTGATTAATAATGGGTAAATATTTTAACTGATGAAGAACTTAGACGCTTC 1451  
 QY 1028 ATAAAAAGATTAATGTTTCAAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087  
 Db 1452 ATAAAAAGATTAATGTTTCAAGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1511  
 QY 1088 TACTGCTCTCTCAG 1147  
 Db 1512 TACTGCTCTCTCAG 1571  
 QY 1148 TTGATTAATCAATCAATTTGCAATTTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1207  
 Db 1572 TTGATTAATCAATCAATTTGCAATTTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631  
 QY 1208 GTGTTAAGCACTAGATTAATGCTGTTGATCTATACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267  
 Db 1632 GTGTTAAGCACTAGATTAATGCTGTTGATCTATACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691  
 QY 1268 ATTCAATTGATTAATAATCTATTTAAAGTTTAAACCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1327  
 Db 1692 ATTCAATTGATTAATAATCTATTTAAAGTTTAAACCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1751  
 QY 1328 CAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1387  
 Db 1752 CAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1811  
 QY 1388 TCACAGCAGACACCCAGATAGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1447  
 Db 1812 TCACAGCAGACACCCAGATAGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1871  
 QY 1448 ATGCTATTAAATAATCTGT 1507  
 Db 1872 ATGCTATTAAATAATCTGT 1931  
 QY 1508 GAACCTTAGCAAAACAGCAAACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1567  
 Db 1932 GAACCTTAGCAAAACAGCAAACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1991  
 QY 1568 ACTGCAAGATGGAG 1627  
 Db 1992 ACTGCAAGATGGAG 2051  
 QY 1628 TTGGCATCGATTTGTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1687  
 Db 2052 TTGGCATCGATTTGTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2111  
 QY 1688 CAGCATCAATGAAG 1747  
 Db 2112 CAGCATCAATGAAG 2171  
 QY 1748 AGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1807  
 Db 2172 AGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2231  
 QY 1808 GCAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1867  
 Db 2232 GCAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2291  
 QY 1868 TGAAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1927  
 Db 2292 TGAAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2351  
 QY 1928 GAGGTTCACACCTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1987  
 Db 2352 GAGGTTCACACCTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2411

QY 1988 TTGCTCACTGAG 2047  
 Db 2412 TTGCTCACTGAG 2471  
 QY 2048 TTACAGCAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2107  
 Db 2472 TTACAGCAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2531  
 QY 2108 TTTCCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2167  
 Db 2532 TTTCCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2591  
 QY 2168 AGGCTTGACCTACAG 2227  
 Db 2592 AGGCTTGACCTACAG 2651  
 QY 2228 CTGCTCTGTCAG 2287  
 Db 2652 CTGCTCTGTCAG 2711  
 QY 2288 AGCCTAATCTGTCAG 2348  
 Db 2712 AGCCTAATCTGTCAG 2772

RESULT 3  
 ADB63315  
 ID ADB63315 standard; cDNA; 2579 BP.  
 XX  
 AC ADB63315;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human cDNA encoding clone TEST120060830.  
 XX  
 KM Human; sex; gene; pharmaceutical; diagnostic; gene therapy;  
 KM tissue regeneration; cell regeneration; membrane protein;  
 KM signal transduction-related protein; transcription-related protein;  
 KM osteoporosis; neurological disease; cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT 54..2543  
 FT CDS /tag= a  
 FT /product= "Clone TEST120060830 protein"  
 XX  
 EP1308459-A2.  
 XX  
 BD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350978.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (RENS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,  
 PI Seki N, Yoshikawa T, Otsuka N, Negahari K, Masuno Y;  
 XX  
 DR WPI: 2003-450961/43.  
 DR P-PDB: ADB65285.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1, Page: 222p; English.



CC The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or  
CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner, and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotide encodes and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumors). The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.

XX Sequence 2579 BP; 797 A; 542 C; 613 G; 627 T; 0 U; 0 Other;

Query Match 94.5%; Score 2200.2; DB 9; Length 2579;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACACAGGCGAAGCTTATGACATGTTAAAGATCAAAAAGGTTCTCAG 67  
DB 319 TTGGTGTGAATCAAGGCGAAGCTTATGACATGTTAAAGATCAAAAAGGTTCTCAG 378  
QY 68 GCATTTAGTAAAGGTTAAGCACTAACATTTCCGGGTGACATCCGCTCCCAAGTGGGCT 127  
DB 379 GCATTTAGTAAAGGTTAAGCACTAACATTTCCGGGTGACATCCGCTCCCAAGTGGGCT 438  
QY 128 TATATCAGTATCACTGATATACCCACTGATGGAAGCCAGAGACTCCCTCAGCTC 187  
DB 439 TATATCAGTATCACTGATATACCCACTGATGGAAGCCAGAGACTCCCTCAGCTC 498  
QY 188 TTCTTTTCAACAGGAAGATCTAATGGAAGGTGATGCTTTTATGGAAGCAATATAT 247  
DB 499 TTCTTTTCAACAGGAAGATCTAATGGAAGGTGATGCTTTTATGGAAGCAATATAT 558  
QY 248 TTTTAACTTAAAGACTACAGCAAAAGTTACTGAAGTTTATGTAAGCCCGAATGAG 307  
DB 559 TTTTAACTTAAAGACTACAGCAAAAGTTACTGAAGTTTATGTAAGCCCGAATGAG 618  
QY 308 AGATGTGAGATACATCACTTTAAACAATGAACTTCACTCAATCAACCAATGTT 367  
DB 619 AGATGTGAGATACATCACTTTAAACAATGAACTTCACTCAATCAACCAATGTT 678  
QY 368 TGCGATCTAATATATATTTTTCAGAGGCTTTTGAAGATCATGATTTGCAAAATG 427  
DB 679 TGCGATCTAATATATATTTTTCAGAGGCTTTTGAAGATCATGATTTGCAAAATG 728  
QY 428 GACGAAATATATATACCCCAATGATGATATTCAGATGCAAGTTTGGTATTT 487  
DB 739 GACGAAATATATATATACCCCAATGATGATATTCAGATGCAAGTTTGGTATTT 798  
QY 488 GGCGTGGCTTACACTTCCATCTCTGATGATGAAACAGCATCATCTCTGACGAG 547  
DB 799 GGCGTGGCTTACACTTCCATCTCTGATGATGAAACAGCATCATCTCTGACGAG 858  
QY 548 TTAGCATTAAGTCTCTGGAAGAGAGCTGTTTGGATTCATGATTTATATATC 607  
DB 859 TTAGCATTAAGTCTCTGGAAGAGAGCTGTTTGGATTCATGATTTATATATC 918  
QY 608 AGACGAAAGACATTAATTTCAAGAGAGTTTCCAAAGACTATATAGTTAGTTTC 667  
DB 919 AGACGAAAGACATTAATTTCAAGAGAGTTTCCAAAGACTATATAGTTAGTTTC 978

QY 668 TTACCAAGTATTAACAATTAAGACATACAGGTGATGATATTTGACTGGGACCAAGATCCCA 727  
DB 979 TTACCAAGTATTAACAATTAAGACATACAGGTGATGATATTTGACTGGGACCAAGATCCCA 1038  
QY 728 AGGCACTTTAAGAAAGCGGACGCTGGAAGTCACTTTAGATATCTACAGGAAGC 787  
DB 1039 AGGCACTTTAAGAAAGCGGACGCTGGAAGTCACTTTAGATATCTACAGGAAGC 1098  
QY 788 AATACAAACCAAGATACACCACTTGAAGCAGCTCTTGTGACGACCCCAAGAA 847  
DB 1099 AATACAAACCAAGATACACCACTTGAAGCAGCTCTTGTGACGACCCCAAGAA 1158  
QY 848 GGGGGGGCTTGGGGGGGACCTGCGAGGGCTGCGAGGCTCATTCCTGAGCTGCTATC 907  
DB 1159 GGGGGGGCTTGGGGGGGACCTGCGAGGGCTGCGAGGCTCATTCCTGAGCTGCTATC 1218  
QY 908 TTACAGGCTTACCTGAATTAATATCGTATGATTTTAACTGATGAAGACTTACCGGTT 967  
DB 1219 TTACAGGCTTACCTGAATTAATATCGTATGATTTTAACTGATGAAGACTTACCGGTT 1278  
QY 968 ATACAGACTTACCTGACAGCAAGGAGGCTGATGAGGAGACATCATTTATCATTC 1027  
DB 1279 ATACAGACTTACCTGACAGCAAGGAGGCTGATGAGGAGACATCATTTATCATTC 1338  
QY 1028 AATAAAGCATTAATGTTCAAGGAGCTTGAAGCTGGGCTTGAAGCTTGAATTCACCT 1087  
DB 1339 AATAAAGCATTAATGTTCAAGGAGCTTGAAGCTGGGCTTGAAGCTTGAATTCACCT 1398  
QY 1088 TACTGCTCTTCTCAGAAAGATTTTGAACAAGAAAGATTCAACAGGTGGAAAAATCAT 1147  
DB 1399 TACTGCTCTTCTCAGAAAGATTTTGAACAAGAAAGATTCAACAGGTGGAAAAATCAT 1458  
QY 1148 TTGATTTACATCCCAATTTGAGATGATGTCGCAAGAAACAAGAGTGACCAATTAATTA 1207  
DB 1459 TTGATTTACATCCCAATTTGAGATGATGTCGCAAGAAACAAGAGTGACCAATTAATTA 1518  
QY 1208 GTGTTAAGCAGCTGATTAATCTGCTGTGATCTATACGGAAGAAATTTAGAACAGCA 1267  
DB 1519 GTGTTAAGCAGCTGATTAATCTGCTGTGATCTATACGGAAGAAATTTAGAACAGCA 1578  
QY 1268 ATTCAATGATCAAAATTTATTTAAAGTTTAAACAAGCATGGGATGCAATGAAAG 1327  
DB 1579 ATTCAATGATCAAAATTTATTTAAAGTTTAAACAAGCATGGGATGCAATGAAAG 1638  
QY 1328 CAATTAATGATGAAGTGAATGACAGAACTGAAGCTTATGAAGTCTTACAGCAAAAG 1387  
DB 1639 CAATTAATGATGAAGTGAATGACAGAACTGAAGCTTATGAAGTCTTACAGCAAAAG 1698  
QY 1388 TCACAGCAGACACCAATAGTGTCTGTCTGTGTCGAAGTATTCGGAAGCAAAATGAG 1447  
DB 1699 TCACAGCAGACACCAATAGTGTGTCTGTCTGTGTCGAAGTATTCGGAAGCAAAATGAG 1758  
QY 1448 ATGCAATTAATAAATTAATCTGTATGACAGATTCCTTACCAAGTCAAGTGTGTGCCCC 1507  
DB 1759 ATGCAATTAATAAATTAATCTGTATGACAGATTCCTTACCAAGTCAAGTGTGTGCCCC 1818  
QY 1508 GAACCTTGAAGCAACAGCAAACTGTATGAGGCTTGTCTTCAAAAGTTCCCTTACAGATGA 1567  
DB 1819 GAACCTTGAAGCAACAGCAAACTGTATGAGGCTTGTCTTCAAAAGTTCCCTTACAGATGA 1878  
QY 1566 ACTGCAAGATGGAAGAGAGGCTCTGAGAGGTGACATCCCTTGAAGCTCGATGATGAG 1627  
DB 1879 ACTGCAAGATGGAAGAGAGGCTCTGAGAGGTGACATCCCTTGAAGCTCGATGATGAG 1938  
QY 1628 TTGGCATGATTTGTTACATGACATGACAGCTGGGCGGAGAGTCAATCGAGAGATTTGTTG 1687  
DB 1939 TTGGCATGATTTGTTACATGACATGACAGCTGGGCGGAGAGTCAATCGAGAGATTTGTTG 1998  
QY 1688 CCAAGCATTAATGAAGGATGACCCGCTGCTTCAACGCTGACATATTTAGATAGAGAC 1747  
DB 1999 CCAAGCATTAATGAAGGATGACCCGCTGCTTCAACGCTGACATATTTAGATAGAGAC 2058

Accession	Gene	Species	Protein	Accession	Gene	Species	Protein
1748	AGAGAGCTGTGATGAGGGGTCTAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	Human	1807	1748	AGAGAGCTGTGATGAGGGGTCTAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	Human	1807
2059	AGAGAGCTGTGATGAGGGGTCTAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	Human	2118	2059	AGAGAGCTGTGATGAGGGGTCTAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	Human	2118
1808	GCAATGAGTACATGCCCCAGCCGGATCATGCTGTGAACCGGATGCGGTAGGAGACGAGCCAGC	Human	1867	1808	GCAATGAGTACATGCCCCAGCCGGATCATGCTGTGAACCGGATGCGGTAGGAGACGAGCCAGC	Human	1867
2119	GCAATGAGTACATGCCCCAGCCGGATCATGCTGTGAACCGGATGCGGTAGGAGACGAGCCAGC	Human	2178	2119	GCAATGAGTACATGCCCCAGCCGGATCATGCTGTGAACCGGATGCGGTAGGAGACGAGCCAGC	Human	2178
1868	TGAAAACACTGCTGTGAACCTAAGAGTGGCCACAGTCTTTGGATTGCTTAATTCATTGGTA	Human	1927	1868	TGAAAACACTGCTGTGAACCTAAGAGTGGCCACAGTCTTTGGATTGCTTAATTCATTGGTA	Human	1927
2179	TGAAAACACTGCTGTGAACCTAAGAGTGGCCACAGTCTTTGGATTGCTTAATTCATTGGTA	Human	2238	2179	TGAAAACACTGCTGTGAACCTAAGAGTGGCCACAGTCTTTGGATTGCTTAATTCATTGGTA	Human	2238
1987	GAGGTTACAACTTGAAGTCTAAGTAAACGTTAATTTGTGTGTAAGAAAAGTGAACACAGATTTT	Human	1987	1987	GAGGTTACAACTTGAAGTCTAAGTAAACGTTAATTTGTGTGTAAGAAAAGTGAACACAGATTTT	Human	1987
2239	GAGGTTACAACTTGAAGTCTAAGTAAACGTTAATTTGTGTGTAAGAAAAGTGAACACAGATTTT	Human	2239	2239	GAGGTTACAACTTGAAGTCTAAGTAAACGTTAATTTGTGTGTAAGAAAAGTGAACACAGATTTT	Human	2239
1988	TTGCTCAAGTCTGAGAGAAAGACTTTCAGATTCACCTCTCTGAAACACTTTTGTATGTAGAG	Human	2047	1988	TTGCTCAAGTCTGAGAGAAAGACTTTCAGATTCACCTCTCTGAAACACTTTTGTATGTAGAG	Human	2047
2239	TTGCTCAAGTCTGAGAGAAAGACTTTCAGATTCACCTCTCTGAAACACTTTTGTATGTAGAG	Human	2358	2239	TTGCTCAAGTCTGAGAGAAAGACTTTCAGATTCACCTCTCTGAAACACTTTTGTATGTAGAG	Human	2358
2048	TTACACAGACCAAGATGTATGACTTTTATGCTGAGCCAGGCTGTGAGAAAGTGTAGTG	Human	2107	2048	TTACACAGACCAAGATGTATGACTTTTATGCTGAGCCAGGCTGTGAGAAAGTGTAGTG	Human	2107
2359	TTACACAGACCAAGATGTATGACTTTTATGCTGAGCCAGGCTGTGAGAAAGTGTAGTG	Human	2418	2359	TTACACAGACCAAGATGTATGACTTTTATGCTGAGCCAGGCTGTGAGAAAGTGTAGTG	Human	2418
2108	TTTCTCCACACATTACAAATGTCAATCTATGACACACAGCGGCTGTAAAGCCAGACCAATAC	Human	2167	2108	TTTCTCCACACATTACAAATGTCAATCTATGACACACAGCGGCTGTAAAGCCAGACCAATAC	Human	2167
2419	TTTCTCCACACATTACAAATGTCAATCTATGACACACAGCGGCTGTAAAGCCAGACCAATAC	Human	2478	2419	TTTCTCCACACATTACAAATGTCAATCTATGACACACAGCGGCTGTAAAGCCAGACCAATAC	Human	2478
2168	AGCGCTTGACCTTACAGAGCTGTGCGCACATCTATTACAGCTGGCCAG	Human	2212	2168	AGCGCTTGACCTTACAGAGCTGTGCGCACATCTATTACAGCTGGCCAG	Human	2212
2479	AGCGCTTGACCTTACAGAGCTGTGCGCACATCTATTACAGCTGGCCAG	Human	2523	2479	AGCGCTTGACCTTACAGAGCTGTGCGCACATCTATTACAGCTGGCCAG	Human	2523

[illegible]

Db	1237	GCATATACACCAAGNATTCACGACCTGGAAGCAGCCGGTCTGGTAGCCAAACCAAGC	1296
Qy	846	AAGCGGGGCTCTGGGGGACATCGCCAGGGCTGCATGCTATTCTTGAGCTCTGCTA	905
Db	1297	GAGGAGAGGCCCCCGCGCACCTCGCCGCGCCAGCTATNTATCTCCCTGAACCTCTGCTA	1356
Qy	906	TCTTAACAGGTCTAATCTGATTAATAATGCGTAATGATTTAAGTAGTAGAACAACCTTAGCCGT	965
Db	1357	TCTCACAGGCTTGACTGTATTAATTCGCATATGATTTCAATGTATGAAGACCTGCGACGT	1416
Qy	966	TCATPACAAGATTAATCTCCAGAGCAAAAGCAGCGTGAAGTGGAGCGACTATGATTAAT	1025
Db	1417	GCACCGCGGTGACCCCTGAGCAGCGGACGGGAAGTGGGCGCGCTCATCGCTACAT	1476
Qy	1026	TCATPAAAACGATTAATGTTCAAAAGGAGCTTCGAGCTGGGGTTTAGGCTTTGATTCAA	1085
Db	1477	CCACAAGGATACAATGTGCAGAGAGAGCTTCGAGCTGGGCGCTGAGCTTCGACTAAA	1536
Qy	1086	CTTACTGCTCTTCCAGGAAATTTTGCAACAAGAAAGATTCACCAAGGTGGAAAAAC	1145
Db	1537	CTTGCTGTCTCTCTGGAAGATCTTATCATCTGAGAAAGATCCACAGGGCGGAAAGAC	1596
Qy	1146	ATTGATTTACATCCACATTTGCAATTTGCAATTTGCTCCAAAGAAACAAGAGTGCACCATTAAT	1205
Db	1597	GTTTATATTACAACCCACAAATTTGGCAGACTGTGTCMAAGAAAACAGAGCGCGCGCTGAT	1656
Qy	1206	TAGTGTAAAGCACTAGATTAATCGGCTGTGTATCTATACGGGAGAAATTAAGACAGC	1265
Db	1657	CAGCTGAAGCCATTGGATTAATCTGCTCTGATCTATTAACCCGACGAAATTAAGAGCAGC	1716
Qy	1266	CAATTCAATTGATACAAATCTATTATAAGTTACACAGCCATGCGCATTCGAATGAGAA	1325
Db	1717	CACTCACTGATPACAGAACTGTTCAAAGTACTCCAGCCATGCGGCAATCCAGATGAAGAAA	1776
Qy	1326	AGCATTAATGATTAAGATGAGATGACAGACTGAAGCCATCTTAAGATCTTACAGCAAA	1385
Db	1777	GGCATCATGATCGAGGTGATGAGACAGAACAGAACTTAATCTGAGAGCTTTCAGCAGAA	1836
Qy	1386	GGTCAACAGACACACCCAGATAGTTGTCTGTCTGTTGTCMAATATGGAAGACAAATA	1445
Db	1837	GGTACGTCAGACACTCAGATAGTTGTCTGTCTCTGTCCTGTCAGATATCGAAGACAAATA	1896
Qy	1446	CGATGCTTTTAAAAATATCTGTGTACAGATTTGCCCTTACCCCAAGTACGTGTGTGGGC	1505
Db	1897	TGATGCCATCAAGAGTACTGTGTACAGACTGGCCCCCAAGTCAAGTGTGTGTGGGC	1956
Qy	1506	CCGAACCTTAAGCAACAGCAAACTGTATGGCCATTGCTCTCAAAAGATTGCCCTACAGAT	1565
Db	1957	CCGAGCCTTGGGCAGACAGCAAAACAGTCAATGGCCATTGCCAACAGATCGGCCAGAT	2016
Qy	1566	GAACTGCAAGATGGAGAGAGAGCTCTGAGGGTGAATCCCCCTGAAGCTCTGTGATGAT	1625
Db	2017	GAACTGCAAGATGGAGAGAGAGCTCTGCGCGGGTGAATGGCCCTGAAACTGGCAATGAT	2076
Qy	1626	CGTTGGCATCGATTTGTTACATGACATGACAGCTGGGGGAGAGTCAATCGACAGATTGT	1685
Db	2077	CGTGGCATCGACTGTATTCATGACACACAGACTGGGGGAGAGTCCATCGCAGATTCGT	2136
Qy	1686	TGCCAGCATCAATGAAGGATGACCCCGTGTTCTCAAGCTGACATATTCAAGATAGAG	1745
Db	2137	CGCCAGCATCAATGAAGGATGACCCCGCTGTTCTCCCGCTGCGCTTTCAGAACCGCGG	2196
Qy	1746	AAGAGAGCTGTAGATGGGCTCAAAAGTCTGCTGGAAGGGCTTCGAGGGCTTGGATAG	1805
Db	2197	GCGAGAGCTGTGTAGATGTCTCAAGGTGTGCTTCAAGCTGCTTCGAGGGCTTGGAGTGG	2256
Qy	1806	CTGCATGAGTACATGCCACAGCCGATCATCGTGTACCGCGATGGCGTGAAGACGACCA	1865
Db	2257	CTGCATGATATCATGCTCCAGCCGTGTCACTGTTACCGAGCGGTGTGGGGAGCGGCA	2316
Qy	1866	GCTGAAAAACATGTGTGAATAGAAAGTGGCACAGTTTTGGATGTCATAAATCCATTGG	1925
Db	2317	GCTGAAGACCTGTCAATTAATGAGGTCCACAGATTCCTAGATGCTCCATGACTGCGG	2376

QY	1926	TAGAGTTACACCCCTTGACTAACGGTAAATTGTGTGTGAAGAAAAGATGAACACCAATT	1985
Db	2377	GAGAGTTACACCCMAAGACTGATCTGTAAATCGTGTAAAGACGTCTCAATGCGAGTT	2436
QY	1986	TTTGTGCTCAGCTGTGAAGAAAGACTTCAGAAATCCACTCTCTGAAACAATATTATGATAGA	2045
Db	2437	TTTGTGCTCAGCTGTGGGGAAGACTTCAGAAACCTCTCTCAGGAGACAATTCATGATGGGA	2496
QY	2046	GATTACGAGACCAAAATGTATGACTTTTTTATCTGAGCCAGGCTGTGAAGTGTAG	2105
Db	2497	AGTCAACGACCAAGATGTGTATGACTTTTTTCATGTGAGCCAGGCAAGTGAAGAACGGGAG	2556
QY	2106	TGTTTCTTCCACACATTTACATGTCACTATGACAAAGGGGCGCTGAAGCAGACCAAT	2165
Db	2557	TGTGTCTCCCAACAACATTCATGTCACTATGACAGAGTGGCTTGAAGCCCGACCAAT	2616
QY	2166	ACAGCGCTTGACCTTACCAAGCTGTGTCCCACTCTATTATTAACACTGCGCAGAGTTCATTCGTGT	2225
Db	2617	CCAAGCGGTGACATACAAAGATCTGCCAAGTGAATATTAATTTGGCTGGAAGTCATCCGAGT	2676
QY	2226	TCTGTGCTCTTGGCAGTACGCGCCCAAGCTGGCTTTTCTTGTGTGGCCAGATATTCCACAG	2285
Db	2677	CCCTGCACTTTGCAAGTATGACACAAAGCTGGCGCTTCTCTGTGGGCCAGAGCATCCACAG	2736
QY	2286	AGAGCGCAATGTCACTGTCAAAACCGGCTTTATACCTCTAA	2328
Db	2737	AGAGCGCAACCTCTCCTGTCTCAACCGGCTTTATACCTCTAA	2779

RESULT 5  
 ID ABV893287 standard; cDNA; 2292 BP.  
 XX ABV893287  
 AC ABV893287;  
 XX  
 DT 13-DEC-2002 (first entry)  
 XX  
 DE Human colon cancer related cDNA SEQ ID NO 2602.  
 XX  
 KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;  
 XX ss.  
 OS Homo sapiens.  
 PN WO200258534-A2.  
 PD 01-AUG-2002.  
 PF 16-NOV-2001; 2001WO-US043704.  
 XX  
 PR 20-NOV-2000; 2000US-0252222P.  
 PR 06-FEB-2001; 2001US-0267011P.  
 PR 28-MAR-2001; 2001US-0279670P.  
 PR 10-JUL-2001; 2001US-0304037P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE,  
 XX  
 DR WPI; 2002-608400/65.  
 DR P-PSDB; ABP67995.  
 XX  
 PT New isolated tumor colon polynucleotide and polypeptide, useful for the  
 PT diagnosis, prevention and/or treatment of cancer, in particular colon  
 PT cancer.  
 XX  
 XX Claim 1; SEQ ID NO 2602; 266pp + Sequence Listing; English.  
 XX  
 CC The invention relates to a human colon tumor expressed polynucleotide  
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of  
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)  
 CC Complements of (i); (iii) at least 20 contiguous residues of (i); (iv)











Db 252 GGGCAGGCTTGTGGACAAACATCCAGAGAAATACCAATGCTCGTTTGAATGAGACCTG 311  
 Qy 1065 GGGTTGAGCTTTGATTCCTTCTCTGAGGAAAGAAATTTTGGCAAGAGAAA 1124  
 Db 312 GGGAGCTGATTTTGGAGCCAGATA--TCTGTAGCTGGCCGAGTTGGCTTGGAGAAA 368  
 Qy 1125 GATTACCAAGGTGAGAAAACATTTGATTAACATCCAAATTTGCAATTTGTCAGAAA 1184  
 Db 369 AATATTATGCAAGACCACTAT--GTCACTGTGTCTGTGTGTGTGTGTGTGTGTGT 425  
 Qy 1185 AACAAAGGTGACCAATTAATGTTTAAAGCCACTAATTAATGCTGTGTGTGTGTGTGT 1244  
 Db 426 TATTGCACTTGTCAAGATTTTAAATGCAAGCTTGTGTGTGTGTGTGTGTGTGTGT 485  
 Qy 1245 GGGAGAAATTAATGAGAGAGCAATTCATGTAACAAATCTTAAAGTTTAAAGTTAC 1304  
 Db 486 CGACAGAACTGAAATGTTTGGAGAGCTTTCTGAACTGTGTGAGAAAGTTGACAGTTC 545  
 Qy 1305 CATGGGCAATGCAAAATGAGAAAAGCAATATGATTTGAGTGTGATGACAGAACTGA 1364  
 Db 546 CATGGGATTTATGAGACTATCCCAAAATCATAAAGTACAGAAATCCAGCTGCAT 605  
 Qy 1365 CTTAAGAGCTTACAGCAAAAGGTGACAGAGACCAAGAAATGTTGTCTGTGTGTGT 1424  
 Db 606 TGTTAGAGCTATACAGCAATATGTTGATCTGTGTGTGTGTGTGTGTGTGTGTGT 665  
 Qy 1425 AAGTATGGAAGAGCAAAATGATGCTTAAATAATCTGTGTGTGTGTGTGTGTGTGT 1484  
 Db 666 TTCTAATGAGAGCTTATATGATTTCAATTAATAATTTGAGCTGAGCTGCCAGT 725  
 Qy 1485 CCAAGTCAAGT 1544  
 Db 726 CCAAGGCAATGT 785  
 Qy 1545 TCAAAAGATTTGCCATACAGATGATGCAAGATGAGAGAGAGCTCTGAGGGGTGACAT 1604  
 Db 786 CACCAAGATGCTATGACAGATGATGCTGAGCTGAGAGAGAGCTGTGGGCTGTGAAAT 845  
 Qy 1605 CCCCCGAAAGCTGT 1664  
 Db 846 ACCTTAAAGTCCGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 905  
 Qy 1665 GAGTCAATGAGAGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1724  
 Db 906 CGTGATGT 965  
 Qy 1725 CTGCATATTTCAAGATGAGAGACAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 1784  
 Db 966 CTGTATCTTTAGAGAAACATGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1025  
 Qy 1785 GGGTCTGAGGGCTTGT 1844  
 Db 1026 AGCAGTCAACAAATGT 1085  
 Qy 1845 CGATGGCTAG 1904  
 Db 1086 TGT 1145  
 Qy 1905 GATTTGTCTAAATTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1964  
 Db 1146 GAGCAGT 1205  
 Qy 1965 GAAAGAGTGAACACAGATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2024  
 Db 1206 GAGAGAGT 1265  
 Qy 2025 TGGAGAGT 2084  
 Db 1266 TGGCAGT 1325  
 Qy 2085 CAGGCTGT 2144  
 Db 1326 CAGGCTGT 1385

Qy 2145 CGGCTGAAGCAGACACATACAGGCTTGAACCAAGCTGTGTGTGTGTGTGTGTGTGT 2204  
 Db 1386 CGGCTTGAAGCCGACATATGACAGAGACTTACATTAATTTGTGTGTGTGTGTGTGT 1445  
 Qy 2205 CTGGCCAGGT 2264  
 Db 1446 CTGGCCGAGT 1505  
 Qy 2265 TGTGTGCAAGTATTCAGAGAGAGCAATGTGTGTGTGTGTGTGTGTGTGTGTGT 2324  
 Db 1506 GGGGCAAGAGATTCATTAAGAACCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1565  
 Qy 2325 CTAA 2328  
 Db 1566 GTGA 1569

RESULT 8  
 ACA62542  
 ID ACA62542 standard; cDNA; 2009 BP.  
 XX  
 AC ACA62542;  
 XX  
 DT 19-AUG-2003 (first entry)  
 XX  
 DE Human full length cDNA encoding Hiwi-Hy protein.  
 XX  
 KW Human; ss; Hiwi-Hy; stem cell maintenance factor; immunosuppressive;  
 KW antidiabetic; antiparkinsonian; cerebroprotective; neuroprotective;  
 KW diabetes; CNS disease; graft-versus-host disease; Parkinson's disease;  
 KW organ transplantation; biological-based sensor; gene.  
 XX  
 OS Homo sapiens.  
 XX  
 FX Key Location/Qualifiers  
 FT CDS 73..1569  
 FT /\*cag= a  
 FT /product= "Hiwi-Hy"  
 FT  
 XX US2003027255-A1.  
 XX  
 XX 06-FEB-2003.  
 XX 26-OCT-2001; 2001US-00029495.  
 XX PE 20-AUG-1999; 99US-00378667.  
 XX FR  
 XX (BALI/) BALINGER D G.  
 XX PA (DRMA/) DRMANAC R T.  
 XX PA (LABA/) LABAT I.  
 XX PA (STAC/) STACHE-CRAIN B.  
 XX PA (DICK/) DICKSON M C.  
 XX PA (JONE/) JONES L W.  
 XX PA (XUEA/) XUE A.  
 XX PA (TANG/) TANG Y T.  
 XX PA (LIUC/) LIU C.  
 XX PA (ASUN/) ASUNDI V.  
 XX  
 XX Ballinger DG, Drmanac RT, Labat I, Stache-Crain B, Dickson MC,  
 XX Jones LM, Xue A, Tang YT, Liu C, Asundi V;  
 XX  
 XX WPI: 2003-479536/45.  
 XX P-PSDB; AB061938.  
 XX  
 XX New isolated stem cell maintenance factor polynucleotide and polypeptide,  
 XX useful for the diagnosis and treatment of disorders such as diabetes, CNS  
 XX diseases, graft versus host disease, Parkinson's disease and organ  
 XX transplantation.  
 XX  
 XX Example 7; Page 48-50; 63pp; English.  
 XX  
 XX The invention relates to a new isolated polynucleotide comprises: (a) the

CC human partial Hiwi-Hy cDNA appearing as AC62541; (b) having the stem  
 CC cell maintenance factor protein coding nucleotide sequence of a  
 CC polynucleotide of (a); or (c) having the mature stem cell maintenance  
 CC factor protein coding nucleotide sequence of (a). Also include are an  
 CC isolated polynucleotide encoding a polypeptide with stem cell maintenance  
 CC factor activity that hybridizes under stringent conditions to the  
 CC complement of (or has a sequence identity greater than 90% to AC62541),  
 CC an expression vector comprising the Hiwi-Hy cDNA, a host cell genetically  
 CC engineered to contain or express the Hiwi-Hy cDNA in operative  
 CC association with a regulatory sequence that controls expression of the  
 CC polynucleotide in the host cell, the partial Hiwi-Hy stem cell  
 CC maintenance factor appearing as AB061932, an anti-Hiwi-Hy antibody,  
 CC detecting Hiwi-Hy in a sample (comprising contacting the sample with a  
 CC compound that binds to and form a complex with the polypeptide),  
 CC identifying the binding compound and modulating cell survival and/or  
 CC death (comprising contacting the cell with Hiwi-Hy for a time sufficient  
 CC to modulate the activity). The methods and compositions of the present  
 CC invention are useful for the diagnosis and treatment of disorders such as  
 CC diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease  
 CC and organ transplantation. They can also be used for re-engineering  
 CC damaged or diseased tissues, manufacture of pharmaceuticals and in the  
 CC development of biological-based sensors. The present sequence is the full  
 CC length cDNA encoding the human Hiwi-Hy protein

XX Sequence 2009 BP; 595 A; 411 C; 458 G; 545 T; 0 U; 0 other;

Query Match 19.7%; Score 458.4; DB 8; Length 2009;

Best Local Similarity 57.2%; Pred. No. 5,3e-126;

Matches 894; Conservative 0; Mismatches 661; Indels 9; Gaps 3;

QY 765 CTCTTGAATTAATACAGAGCAATACCAAGATACCGACTTGAAGAGCTGT 824  
 DB 15 CTAGTGGATTACTACAGAGCGAGTATGATTACTGTATCGAAGCTGATAGCCCAT 74  
 QY 825 CTGTGTACGACGAGCCAGAGAGAGCGGGGCGGCGGAGCACTGCGAGGCTGCAT 884  
 DB 75 GCTTGTATGTCTTTAAG--AAGAAAGAAATGACACAGTGAAGCTTACGCTCCCA 131  
 QY 885 GCTCATTCCTGAGCTCTGCTATCTTACAGGCTTAACTGATTAATGCTATGATTTAA 944  
 DB 132 CCGTATACCTGAGCTGCTCTTTTACAGGGCTGACTGACAGGCAACATCTGATTTCCA 191  
 QY 945 CGTATGAAAACATTAGCCGCTTCAACAGCTAACCCAGAGAAAGGAGGCTGAGT 1004  
 DB 192 GCTATGAGGCTGTGCTGAAAAGACGCTGACGCTTCAAGCCGAGAGCGGCT 251  
 QY 1005 GGAAGCACTGATTGATTCATTCAATAAAGATATGTTCAAGAGGAGCTTCGAGACTG 1064  
 DB 252 GAGCAGGCTTGTGACACATCCAGAGAAATACCAATGCTCGCTTGAATAGAGACTG 311  
 QY 1065 GGGTTGAGCTTGTGCTCAACTTACTGCTCTCAAGAGAAATTTTGGCAACAGAAA 1124  
 DB 312 GGAATGATTTTGGAGCCAGATA--TCTCTGACGCGGAGTGTGCTTCAAGAAA 368  
 QY 1125 GATTACCAAGGTGAAAACATTGATTAACAATCAATTCAGATTGATGTCACAAAGA 1184  
 DB 369 AATATTAATGCAAGACCAAT--GTCAACCTGTGTCTGTGTGATCTGTCACAGGA 425  
 QY 1185 AACCAAGAGTGCACATTATATGTTTGAAGCCACTGATTAATGCTGTGATCTATAC 1244  
 DB 426 TATTCGAAGCTTCAAGATTTTAAATGACAGTCTTGAATACCTGTGATTTATGAG 485  
 QY 1245 GCGAAGAAATTAAGAGCGCAATTCATTGATTAACAATCATTTAAGTTACACAGC 1304  
 DB 486 CGACGAAGTGAATGTGTCGAGAGCTTTCTGAAGCTTGAAGAGTTGACAGTTC 545  
 QY 1305 CATGGCATGCAATGAGAAAAGCAATATGATTAAGTGAAGTGAAGTGAAGCTTA 1364  
 DB 546 CATGGATTTAATGTGATCAACCCCAATCAATAAAGTCAAGAAATTCACAGCTGCA 605  
 QY 1365 CTTAAGAGCTTACAGCAAAAGTCAACAGACACCCAGATGTTGTCTGTCTGTGTC 1424  
 DB 606 TGTTAGAGCTATACAGCAATATGTTGATCTGATGTTGCTGATATGCTATCTGCC 665

QY 1425 AAGTATCGAAGAGCAATACAGATGCTATTAATAAATACCTGTGACAGATTGCCCTAC 1484  
 DB 666 TTCTATACAGAAAGCTTATATGATTCATTAATAAATATTTGAGCTCAACAGCTCCAGT 725  
 QY 1485 CCCAAGTCAAGTGTGTGTGCGGCAACCTTAAGGCAACACGAAACGTGATGCGCATTCG 1544  
 DB 726 CCCAAGCAATGTGTGTGTCTGCGACTTGAATTAACAGGGCATGATGATGATTCGC 765  
 QY 1545 TACAAGTGTCCCTACAGATGAAGTCAAGATGAGAGAGCTGAGGTTGACAT 1604  
 DB 786 CACCAAGTGTCTATGAGATGATGATTCAGAGCTGAGAGGAGCTGTGAGTGAAT 845  
 QY 1605 CCCCCTGAAGCTGTGATGATGCTGGCATGATGATTTTACATGACATGACAGTGGGCG 1664  
 DB 846 ACCTTTAAAGTCCCTGATGTGTGTGCGATTTGAATGATGATGATGATGATGATGAT 905  
 QY 1665 GAGTCAATCGCAGATTTGTCAGACATGATGAAGGATGACCCGCTGTCTGACG 1724  
 DB 906 CGTATGTGTTGTGATGCGTGGCCAGTGTTAACCCAGAAATACAGAGTGTTCGCG 965  
 QY 1725 CTGCATATTTGAGATGAGAGACAGAGCTGTGATGAGTGTCTAAAGTCTGCTGCAAGC 1784  
 DB 966 CTGATCTCTTACAGAAACATGACTGATGATGATGATGATGATGATGATGATGATGAT 1025  
 QY 1785 GAGCTTGAGGCTTGAATTAAGTCAATGATGATGATGATGATGATGATGATGATGATGAT 1844  
 DB 1026 AGCACTCAACAAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1085  
 QY 1845 CGATGCGCTGAGAGAGCGGCGAGCTGAAAACATGCTGATGATGATGATGATGATGATGAT 1904  
 DB 1086 TGCTGTGATGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1145  
 QY 1905 GATTTGTCAAAATTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1964  
 DB 1146 GAGCAGTGTGAGAAATCAAGTCAATTAACAGCTCAAGATGATGATGATGATGATGATGAT 1205  
 QY 1965 GAAAGAGTGAACACCAATTTTGTCTCACTGTGAGAAAGATTCAGATTCAGCTTCC 2024  
 DB 1206 GAGAGAGTGAATGCAACATTTTACCGAAATGAACCGACTGTACAGAAACCCCACT 1265  
 QY 2025 TGAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2084  
 DB 1266 TGGCACTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1325  
 QY 2085 CCAAGTGTGAGAGAGT 2144  
 DB 1326 CCAAGTGTGAGAGAGT 1385  
 QY 2145 CGGCTGAAAGCAGACACATACAGGCTTGAAGCTTCAAGCTGTGCTCATTTAGCA 2204  
 DB 1386 CGGCTGAAAGCAGACACATACAGGCTTGAAGCTTCAAGCTGTGCTCATTTAGCA 1445  
 QY 2205 CTGGCCAGAGTCAATTCGT 2264  
 DB 1446 CTGGCCAGAGTCAATTCGT 1505  
 QY 2265 TGTGTGCAAGATTTCAAGAGAGCAATTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2324  
 DB 1506 GGGGAGCAAAAGATTCATTAAGAACCCAGTGTGATTAAGCCACATCTCTTACCT 1565  
 QY 2325 CTAA 2328  
 DB 1566 GTGA 1569

## RESULT 9

ADA53334 ID ADA53334 standard; cDNA; 3620 BP.

ADA53334;

20-NOV-2003 (first entry)



QY 1611 GAAGCTGTGATGATCGTTGGCATCGATTTGTACATGACATGACAGCTGGGCGAGATC 1670  
 DB 2340 GAAACAGTTAATGTTGATCGGAGATGATGTTTACATGACCCCGATGAGGATCGCTC 2399  
 QY 1671 AATGCGAGATTTGTTGCGACATCAATGAAGGATGACCGCTGGTTCTTCAGCTGCAT 1730  
 DB 2400 CGTGGTGGCTTCGCGACATCAATCTCACCTTCACAAATGATATCCCGGATGCT 2459  
 QY 1731 ATTTCAGATAGAGACAGAGAGCTGTGATGATGGGCTCAAGTCCCTGACAGGAGGCTC 1790  
 DB 2460 GTTCCAGATGCCGATCGAGAGATGTGACAGCCTGAGCTAATGCTTCGTGGGCTCTT 2519  
 QY 1791 GAGGCGTTGATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1850  
 DB 2520 AAAAAAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2579  
 QY 1851 CGTAGAGAGCG 1910  
 DB 2580 AGTGTCTGATGCGCACTGAAAGACGTTGCCAATGATGATGATGATGATGATGATGATG 2639  
 QY 1911 TCTAAATCCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1970  
 DB 2640 TTTTGAAGCTTTTGAGAA---TTATGACCCCAAGATGATGATGATGATGATGATGATG 2696  
 QY 1971 AGTGAACACGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2030  
 DB 2697 AATCAGTACTATCTATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2756  
 QY 2031 AGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2090  
 DB 2757 TGTGTATATCATACATACATACATACATACATACATACATACATACATACATACATAC 2816  
 QY 2091 TGTGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2150  
 DB 2817 TGTACGAGGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2876  
 QY 2151 GAGGCGAC 2210  
 DB 2877 GAGCCTGATCATATGACAGAGCTGATCTTCAAACTGTGACATGATGATGATGATGATG 2936  
 QY 2211 AGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2270  
 DB 2937 TGGCACCATCAGAGATTCAGAGCTCCTTGCAGATGATGCCACAGAGTATGCTTCTGTG 2996  
 QY 2271 CCAGAGTATTCACAGAGAGCCAAATCTGTCACTGTCAACCGCCTTACTACTCTAA 2328  
 DB 2997 ACACATCTTGCATCATGACACAGCATCAGCTGTGAGAACTGTTCTTCTGTGA 3054

RESULT 10  
 AAF21880  
 ID AAF21880 standard; DNA; 423 BP.  
 AC AAF21880;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 267.  
 XX  
 KM Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KM neurotropic; neurprotective; antiviral; antiallergic; hepatocarcinoma;  
 KM antidiabetic; antineoplastic; antitumor; antitumor; antitumor;  
 KM antibacterial; antifungal; antiparasitic; cardiac; immune disorder;  
 KM Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KM cardiovascular disorder; wound healing; neurological disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055173-A1.  
 XX  
 PD 21-SEP-2000.

XX  
 PF 08-MAR-2000; 2000MO-US005881.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-611515/58.  
 DR P-PSDB; AAB58977.  
 XX  
 PT New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention, treatment  
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and  
 PT neurological diseases.  
 XX  
 PS Claim 1; Page 686; 1299pp; English.  
 CC  
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic;  
 CC neuroprotective; antiviral; antiallergic; hepatocarcinoma; antidiabetic;  
 CC antineoplastic; antitumor; antitumor; antitumor; antitumor;  
 CC antifungal; antiparasitic and cardiac activity. The polynucleotide and  
 CC protein sequences are used in the diagnosis of cancer, particularly  
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists  
 CC and agonists may also be used in the diagnosis, prevention and treatment  
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC cardiovascular disorders such as myocardial ischaemia; wound healing;  
 CC neurological diseases such as cerebral anoxia and epilepsy; and  
 CC infectious diseases  
 XX  
 SQ Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 0 U; 4 Other;  
 Query Match 17.0%; Score 395.4; DB 3; Length 423;  
 Best Local Similarity 98.8%; Pred. No. 1.6e-107;  
 Matches 404; Conservative 4; Mismatches 0; Indels 1; Gaps 1;  
 QY 1474 GATTGCTTACCCCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1533  
 DB 14 GATTGCTTACCCCAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 73  
 QY 1534 ATGGCCATTGCTCAAAAGTTGCTCTCAAGATGATGATGATGATGATGATGATGATGATG 1593  
 DB 74 ATGGCCATTGCTCAAAAGTTGCTCTCAAGATGATGATGATGATGATGATGATGATGATG 133  
 QY 1594 AGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1653  
 DB 134 AGGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 193  
 QY 1654 ACAGCTGGGCGAGAGTCAATGCGAGAGTTGTTGCCAGATCATGATGATGATGATGATG 1713  
 DB 194 ACAGCTGGGCGAGAGTCAATGCGAGAGTTGTTGCCAGATCATGATGATGATGATGATG 253  
 QY 1714 TGGTTCACGCTGCAATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1773  
 DB 254 TGGTTCACGCTGCAATTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATG 313  
 QY 1774 TGGCTCAAGCGGCTGTGAGGCTTGGATGATGATGATGATGATGATGATGATGATGATG 1833  
 DB 314 TGGCTCAAGCGGCTGTGAGGCTTGGATGATGATGATGATGATGATGATGATGATGATG 373  
 QY 1834 ATGCTGATCGGAGTGGCGTGAAGAGCGCGAGTGAATGATGATGATGATGATGATGATG 1882  
 DB 374 ATGCTGATCGGAGTGGCGTGAAGAGCGCGAGTGAATGATGATGATGATGATGATGATG 421

RESULT 11

AA13630 standard; CDNA; 3649 BP.

AA13630;

18-DEC-2001 (first entry)

CDNA sequence encoding mammalian Spgl6.

Mammalian; reproductive-specific protein; male infertility; gene therapy; spermatogenesis; sperm count disorder; anti infertility; reproduction; ss.

Mammalia.

WO20016752-A2.

13-SEP-2001.

07-MAR-2001; 2001WO-US007371.

07-MAR-2000; 2000US-0187518P.

12-JAN-2001; 2001US-0261557P.

(WHEB) WHITEHEAD INST BIOMEDICAL RES.

Wang PJ, Page DC;

WPI: 2001-570774/64.

P-PDB; AA007866.

Novel reproduction-specific protein, useful for treating disorders of reduced sperm count, enhancing/increasing sperm count and/or sperm activity.

Claim 1; Fig 15; 151pp; English.

The present invention relates to the isolation of novel mammalian and human reproductive-specific proteins (AA007859-AA007899), and the nucleic acids encoding them. The nucleic acids encoding reproductive-specific proteins are useful for diagnosing infertility which is a result of reduced sperm count, reduced sperm motility, malformed sperm or combinations of these. The sequences of the invention are useful as markers for spermatogenic cells, for identifying genes or proteins characteristic of male infertility, diagnosing or aiding in the diagnosis of infertility in men, and for contraception in which sperm production or sperm count is reduced or defective sperm is produced. Antibodies to reproductive-specific proteins are useful for determining the presence of these proteins in a sample obtained from a man being assessed for infertility, for identifying the expression of genes in particular cell type or particular developmental stage, for studies of spermatogenesis, and for immunolocalization of germ cells or in Western blots for assessing the presence of the protein the antibody binds. The sequences of the invention are also useful for treating disorders of reduced sperm count, and for increasing sperm count and/or sperm activity. The nucleic acids of the invention are useful in gene therapy. AA13623-AA13647 represent cDNA sequences encoding for the mammalian reproduction-specific proteins of the present invention

Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 U; 0 Other;

Query Match 16.4%; Score 381; DB 5; Length 3649;

Best Local Similarity 52.7%; Pred. No. 1.2e-102; Matches 975; Conservative 0; Mismatches 855; Indels 21; Gaps 6;

QY 490 CTTGGCTTCACTCTTCATCTTCATGTAAGAAACAGATCATGCTGCAGCTGAGCT 549  
DB 1 CTTGGCTTATCGGCTGATGATCCGAGACAGACGGGGTCTTCTCTGCTGCTATATGTC 60  
QY 550 AGCCATAAGTCTCTGCAAGTGAAGACTGTTTGGATTTCATGTTCACTTTATCATGAG 609

DB 61 TCTCATAGGTCACTTCGGAACGACTGTGTGATGATGATCATGCTATCTACCGAG 120  
QY 610 ACAGAAAGCATTAATTTCCAAAGCAAGTTTCCAAAGCACTAATAGTTTGTGTTCT 669  
DB 121 AACAGAGAGCAGC---TTCCAGAGACGATGCAAGCAAGCTTGTGGAGCAGCTTATC 177  
QY 670 ACCAGTATTAATTAAGACATACAGAGGATGATATGATCTGGAGACCAAGTCCAG 729  
DB 178 ACAGCTTACAGATGATGATCTTACCGATGATGATGATGATGATGATGATGATGAT 237  
QY 730 AGCAGCTTTAAGAAAGCCAGGCTTGAAGTCACTTTGATATCTACAGAGAGCA 789  
DB 238 GACAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 297  
QY 790 TACACCAAGAGATCCAGACTTGAAGAGAGCTGTCTGTGGAGAGAGCCCA---AGAG 846  
DB 298 TATGGATACAGTCAAGAGAGATGACAGCGCTGTGATCAACCGCCAGTAGAGAG 357  
QY 847 AGGCGGGGCTTGGGGGAGCACTGCCAGGAGCTGCCAGCTTCTGATGATGATGAT 906  
DB 358 CAGATTAACATGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 417  
QY 907 CTACAGCTTCACTGATTAATATGCTATGATTTAAGTATGATGATGATGATGATGAT 966  
DB 418 ATGACGGGATCTCTGAGAAAGATGAAGAGACTTCAAGCCATGAGAGCTTGAATCAG 477  
QY 967 CATAAGAGTAACTCCAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026  
DB 478 CAGATTAACATGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
QY 1027 CATTAAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086  
DB 538 TCACAAAG 597  
QY 1087 TTAAGTCTTCTCAG 1146  
DB 598 GTCCACAGATTAAG 657  
QY 1147 TTTGATTAACATGAG 1206  
DB 658 TTTGAT---CAGATGAG 714  
QY 1207 AGGTTAAG 1266  
DB 715 CTAATCTTCCATGATCTTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774  
QY 1267 AATTCAATGATTAAG 1326  
DB 775 AGAGATCTGATTAAG 834  
QY 1327 GCATTAATGATTAAG 1383  
DB 835 CAGAGCTGAG 894  
QY 1384 AAGGTCAAG 1447  
DB 895 TTAATGAG 954  
QY 1438 GACAAATGAG 1497  
DB 955 GATCTTAATGAG 1014  
QY 1498 GTGGTGGCCGAG 1557  
DB 1015 ATCAATGAG 1074  
QY 1558 CTAAGATGAATCTGAG 1617  
DB 1075 CTTCAAGATGAATCTGAG 1134  
QY 1618 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677  
DB 1135 CTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1194

QY	1678	GGATTGGTTGCCAGCATTCATGAAAGGATGACCCCGCTGGTCTTCACAGCTGCATTAATTGCG	1737
Db	1195	GGCTTGCGTGCCAGCATTAATATCTCACTCACCAAAATGGTACTGAGGGTGGTCTCCAG	1254
QY	1738	GATAGAGSACAGAGAGCTGGTAGATGGCTCCAAAGTCTGCCTGCAGACGGCTCTGAGGGCT	1797
Db	1255	ATGCCACATCAGGAGATTGTGGACAGGCTGAACTCTGCGCTGGTGGGTTCTTGAAAAAG	1314
QY	1798	TGGAAATGCTGCMAATGATGATCAATGCCACGGCGGATCATCGTGTACCGCGATGGCGTAGGA	1857
Db	1315	TATTATGAGGTGAHCCATTGTCTCCAGAGAAATTTGGTGTGTAACCGAATGAGATGTCT	1374
QY	1858	GACGGCCAGCTGAAAACACTGGTGAACCTAGAGGTGCAAGTTTGGATTGTCTTAAA	1917
Db	1375	GATGGCCAGCTAAAAGACATGGCCCACTAGAGATCTCCTCAGCTGCAGAAAGTGTTTGAA	1434
QY	1918	TCCATTGTGAGGTATCAACCCTAGACTAAACGGTAATTGGTGGAAAGAAAGAGTGAAC	1977
Db	1435	GCCTTTGATA--ACTHCAHCCCAAAATGGTGGTGTGGTAGTTGAGTAAGAAAAATCAGC	1491
QY	1978	ACCAGAATTTTTTGCTCAAGTCTGGAGGAAAGACTTAGAATCCACTTCTGTGAAAGACTATT	2037
Db	1492	ACCAATCTGTACCTTGCTGCTCTGTATCACTTCTGTAAACCCCTCCCGGAGCTGTGTT	1551
QY	2038	GATGTAGAAGTTTACACAGACCAAGATGGTATGACTTTTATTCGTAGGCCAGGCTGTAGA	2097
Db	1552	GATCATACCAATAACCAAGCTGTGAGTGGGTGATTTCATCCTTCTTGCCCATCAATGTGCA	1611
QY	2098	AGTGTATGTTTTCTCCCAACAATTACATATCATCTATGACCAAGCGGCTGAAAGCA	2157
Db	1612	CAGGCTGTGGCAATCTACACACTACACTCTGTGTTCTGAAACATGCAAAATCTBAGCCT	1671
QY	2158	GACCACATACAGCGCTTGACCTAACAGCTGTGCAATCTATTACAACTGGCCAGGTGTC	2217
Db	1672	GATCACATGACAGAGGTGATCTTTCAACTATGCCCCACATGTACTGGAAATGGCCGTGTACC	1731
QY	2218	ATTGCTGTTCTCGCTCCTTGCCAGTAAACCCCAACAGCTGGGCTTTCTTGTTGGCCAGAGT	2277
Db	1732	ATCCGAGTTCCAGCTCTTTCGACAGTAATGCCCAAGATGAGCTTCTGTCCGGACAGATT	1791
QY	2278	ATTCCACAGAGACCAATCTGTCACTGTCCAAACCCCTTTACTACTCTTAA	2328
Db	1792	TTGCATCATGAGCCAGCATTCAGCTGTGTGGGAACTCTGTTCTCTGTAA	1842

RESULT 12	
AB219511	
ID	AB219511 standard, cDNA, 367 BP.
XX	
AC	AB219511;
XX	
DT	23-JAN-2003 (first entry)
DE	Group III cDNA cancer related clone SEQ ID NO:1937.
XX	
KW	Human; cancer; tumour; therapy; diagnosis; CT antigen; Cp antigen;
KW	immune response; virology; immunology; microbiology; molecular biology
KW	recombinant DNA technology; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200278516-A2.
XX	
PD	10-OCT-2002.
XX	
PF	28-MAR-2002; 2002MO-US010421.
XX	
PR	30-MAR-2001; 2001US-0280255P.
PR	28-AUG-2001; 2001US-0315563P.
XX	
PR	09-JAN-2002; 2002US-0347313P.
XX	
PA	(CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Galger A;  
XX WPI: 2003-058387/05.  
XX  
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,  
XX preventing and treating cancer expressing CT or CP mRNA antigens, and in  
XX virology, immunology, microbiology, molecular biology and recombinant DNA  
XX techniques.  
XX  
XX Claim 1; SEQ ID NO 1937; 207pp; English.  
XX  
XX ABO17575 to ABO20506 represent isolated polynucleotide (I) sequences, and  
XX ABO54446 to ABO54472 represent protein (II) sequences, from the present  
XX invention. (I) and (II) have cytostatic activity and can be used in gene  
XX therapy and vaccines. (I), (II), antibodies and compositions from the  
XX present invention are useful for diagnosing, preventing and treating  
XX cancer, which expresses CT or CP mRNA antigens. They are useful for  
XX stimulating immune response. They can also be useful in virology,  
XX immunology, microbiology, molecular biology and recombinant DNA  
XX techniques. N.B. The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from Wipo at [ftp.wipo.int/pubd/published\\_pct\\_sequences](http://wipo.int/pubd/published_pct_sequences)  
XX  
XX Sequence 367 Bp; 118 A; 77 C; 91 G; 81 T; 0 U; 0 Other;

Query Match	15.7%	Score 365.4	DB 7	Length 367
Best Local Similarity	99.7%	Pred. No. 1,4e-96		
Matches 366	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1291	AAAGTTACACGAGCATGGGCGATGCAATGATGAGAAAAGCATTAATGATTGAATGATGATGAC	1350	
DB	1	AAAGTTACACGAGCATGGGCGATGCAATGATGAGAAAAGCATTAATGATTGAATGATGATGAC	60	
QY	1351	AGAACTGAAGCCTTACTTAAGATCTTACAGCAAAAGTCAACGAGACACCCAGATGTT	1410	
DB	61	AGAACTGAAGCCTTACTTAAGATCTTACAGCAAAAGTCAACGAGACACCCAGATGTT	120	
QY	1411	GTCTCTCTGTTGTCAAGTATTCGGAAGGACAAATACATGCTATTAAAAAATACCTGTGT	1470	
DB	121	GTCTCTCTGTTGTCAAGTATTCGGAAGGACAAATACATGCTATTAAAAAATACCTGTGT	180	
QY	1471	ACAGATTGCCCTTACCCCAAGTCACTGTGTGTGTGGCCCGAACCCTTAGCGAAACACCAACT	1530	
DB	181	ACAGATTGCCCTTACCCCAAGTCACTGTGTGTGTGGCCCGAACCCTTAGCGAAACACCAACT	240	
QY	1531	GTCAATGGCCATGTGTACAAAGATTGCCCTTACAGATGATCTGTAATCGAAGTGGAGAGAGCTC	1590	
DB	241	GTCAATGGCCATGTGTACAAAGATTGCCCTTACAGATGATCTGTAATCGAAGTGGAGAGAGCTC	300	
QY	1591	TGAGAGGTGACATCTCCCTGAAAGCTGCTGATGATTCGTGGCCATCGAATGTTCACATGAC	1650	
DB	301	TGAGAGGTGACATCTCCCTGAAAGCTGCTGATGATTCGTGGCCATCGAATGTTCACATGAC	360	
QY	1651	ATGACAG 1657		
DB	361	ATGACAG 367		
RESULT 13				
XX	ABZ2005	ABZ2005 standard; cDNA; 367 BP.		
XX	ABZ20005			
XX	DT	23-JAN-2003 (first entry)		
XX	DE	Group III cDNA cancer related clone SEQ ID NO:2431.		
XX	KX	Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;		
XX	KW	immune response; virology; immunology; microbiology; molecular biology;		
XX		recombinant DNA technology; gene; ss.		
XX				



OS Homo sapiens.  
 XX WO200278516-A2.  
 XX  
 XX 10-OCT-2002.  
 XX  
 XX 28-MAR-2002; 2002WO-US010421.  
 XX  
 XX 30-MAR-2001; 2001US-0280255P.  
 XX 28-AUG-2001; 2001US-031563P.  
 XX 09-JUN-2002; 2002US-0347313P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Wang T, Wang S, Bangur CS, Gaiger A;  
 XX WPI, 2003-058387/05.  
 XX  
 XX New immunogenic polynucleotides or polypeptides useful for diagnosing,  
 XX preventing and treating cancer expressing CT or CP mRNA antigens, and in  
 XX virology, immunology, microbiology, molecular biology and recombinant DNA  
 XX techniques.  
 XX  
 XX Claim 1; SEQ ID NO 2431; 207pp; English.  
 XX  
 XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and  
 XX ABP54446 to ABP54472 represent protein (II) sequences, from the present  
 XX invention. (I) and (II) have cytostatic activity and can be used in gene  
 XX therapy and vaccines. (I), (II), antibodies and compositions from the  
 XX present invention are useful for diagnosing, preventing and treating  
 XX cancer, which expresses CT or CP mRNA antigens. They are useful for  
 XX stimulating immune response. They can also be useful in virology,  
 XX immunology, microbiology, molecular biology and recombinant DNA  
 XX techniques. N.B. The sequence data for this patent did not form part of  
 XX the printed specification, but was obtained in electronic format directly  
 XX from WFO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 367 BP; 118 A; 78 C; 89 G; 79 T; 0 U; 3 Other;  
 XX  
 XX Query Match 15.6%; Score 364; DB 7; Length 367;  
 XX Best Local Similarity 99.2%; Pred. No. 3.7e-98;  
 XX Matches 364; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 XX  
 XX 1291 AAGATTACCCGCGCATGGGCAATGAGAAAGCAATATGATGAGTGATGAC 1350  
 XX 1 AAGATTACCCGCGCATGGGCAATGAGAAAGCAATATGATGATGATGATGAC 60  
 XX  
 XX 1351 AGAAGTGAAGCTTAAAGAGCTTACAGCAAAAGGTGACAGCAGACCCAGATAGTT 1410  
 XX 61 AGAAGTGAAGCTTAAAGAGCTTACAGCAAAAGGTGACAGCAGACCCAGATAGTT 120  
 XX  
 XX 1411 GTCTGTCTGTTCAAGTATTCGAAAGCAATATGAGCTTAAATATACCTGTGT 1470  
 XX 121 GTCTGTCTGTTCAAGTATTCGAAAGCAATATGAGCTTAAATATACCTGTGT 180  
 XX  
 XX 1471 ACAGATGCGCTTACCCGCAAGTCAAGTGTGTGGTGGCCGCAACCTTAGGCAAAAGCAAACT 1530  
 XX 181 ACAGATGCGCTTACCCGCAAGTCAAGTGTGTGGTGGCCGCAACCTTAGGCAAAAGCAAACT 240  
 XX  
 XX 1531 GTCATGGCCATTGCTCAAAAGATTTCCCTTCAATGATGACGAATGGAGAGAGAGCTC 1590  
 XX 241 GTCATGGCCATTGCTCAAAAGATTTCCCTTCAATGATGACGAATGGAGAGAGAGCTC 300  
 XX  
 XX 1591 TGGAGGGTGGACATCCCGCTGAAGCTGTGATGATGCTTGGCATTCATTTACATGAC 1650  
 XX 301 TGGAGGGTGGACATCCCGCTGAAGCTGTGATGATGCTTGGCATTCATTTACATGAC 360  
 XX  
 XX 1651 ATGACAG 1657  
 XX 361 ATGACAG 367  
 XX  
 XX Db  
 XX  
 XX RESULT 14

AAH15960  
 XX ID AAH15960 standard; cDNA; 2272 BP.  
 XX  
 XX AAH15960;  
 XX  
 XX 26-JUN-2001 (first entry)  
 XX  
 XX Human cDNA sequence SEQ ID NO:14558.  
 XX  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX EPI074617-A2.  
 XX  
 XX 07-FEB-2001.  
 XX  
 XX 28-JUL-2000; 2000EP-00116126.  
 XX  
 XX 29-JUL-1999; 99JP-00248036.  
 XX 27-AUG-1999; 99JP-00300253.  
 XX 11-JAN-2000; 2000JP-00118776.  
 XX 02-MAY-2000; 2000JP-00183767.  
 XX 09-JUN-2000; 2000JP-00241899.  
 XX  
 XX (HELI-) HELIX RES INST.  
 XX  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI, 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 XX length cDNAs defined in the specification, and for the detection and/or  
 XX diagnosis of the abnormality of the proteins encoded by the full-length  
 XX cDNAs.  
 XX  
 XX Claim 8; SEQ ID NO 14558; 2537pp + Sequence listing; English.  
 XX  
 XX The present invention describes primer sets for synthesizing 5602 full-  
 XX length cDNAs defined in the specification. Where a primer set comprises:  
 XX (a) an oligo-dT primer and an oligonucleotide complementary to the  
 XX complementary strand of a polynucleotide which comprises one of the 5602  
 XX nucleotide sequences defined in the specification, where the  
 XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 XX of an oligonucleotide comprising a sequence complementary to the  
 XX complementary strand of a polynucleotide which comprises a 5'-end  
 XX sequence and an oligonucleotide comprising a sequence complementary to a  
 XX polynucleotide which comprises a 3'-end sequence, where the  
 XX oligonucleotide comprises at least 15 nucleotides and the combination of  
 XX the 5'-end sequence/3'-end sequence is selected from those defined in the  
 XX specification. The primer sets can be used in antisense therapy and in  
 XX gene therapy. The primers are useful for synthesizing polynucleotides,  
 XX particularly full-length cDNAs. The primers are also useful for the  
 XX detection and/or diagnosis of the abnormality of the proteins encoded by  
 XX the full-length cDNAs. The primers allow obtaining of the full-length  
 XX cDNAs easily without any specialised methods. AAH03166 to AAH13629 and  
 XX AAH13630 to AAH18742 represent human cDNA sequences; AAH2446 to AAH95893  
 XX represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 XX oligonucleotides, all of which are used in the exemplification of the  
 XX present invention  
 XX  
 XX Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 U; 0 Other;  
 XX  
 XX Query Match 14.2%; Score 330; DB 4; Length 2272;  
 XX Best Local Similarity 51.6%; Pred. No. 1.8e-87;  
 XX Matches 884; Conservative 0; Mismatches 810; Indels 18; Gaps 5;  
 XX  
 XX 629 AAGAAAGAGTTCCAAAGAACTAATAGTTAGTTGTTCTTACCAAGATATACATAGA 688  
 XX 1 AGAATGAAGTGTACTAAGCTTCTGTTGGCAATATTTATATCACTGATATTAACAATCTGA 60  
 XX  
 XX 689 CATACAGAGTGATGATATTAAGCTGGAACCAAGATCCCAAGAGCACTTTAAGAAAGCCG 748

Db	61	CCATCGTATTGATGATGGATTGGAAATTAACATCCAAAGGATAGCTTCAAGATGCTG	120
Qy	749	ACGGCTCTAAGTCAAGCTTCTTAGAATACTACAGAAAGCAATACCAAGATCAACG	808
Db	121	ATGGAAAGAGATCAATCTTGGAATACTACAGCAAAATATAGGATCAAGTTAAG	180
Qy	809	ACTTGAAGCAAGCTGCTTGTGTGACGACCCCA--AAGAAAGGGGGGCCCTGGGGGGA	865
Db	181	AAGAGAGCCAGCCATTGCTGATTCAGAGGCCAGTAGAGAGACAGATTAATCTGGATGC	240
Qy	866	CAGTGCAGGGCCCTGCATGCTCATTCCTGAGGCTGCTGCTATCTTCAAGTCTTACTGATA	925
Db	241	TGCTTAAAGGGGAAATCTCTGCTGCTGCTGAGGTTTCTTTATGACCGGAATCCAGAGA	300
Qy	926	AAATGCGTAAATGATTTTAACTGATGAAGAATTACCGCTTATACAGACTPACTCAG	985
Db	301	AGATGAAGAAGGACTTCAGAGCCATGAAGGATTTGGCTCAGCAATCAATCTGAGCCCA	360
Qy	986	AGCAAAAGCAGCGGTGAAGTGGAGCAACTCATGATTACATTATPAAAAAGATAATGTTTC	1045
Db	361	AGCAACCAATAGTCTTGGAAATGCTGTGTCGAAGAAATGGCAAGAACGAGAGCCCA	420
Qy	1046	AAAGGAGCTTGCAGACTGAGGTTTGAGCTTTGATTCCACTTACTGTCTCTCAGGA	1105
Db	421	CCATGAACATGATCGCTGGGGGGCTCGCTGTCGAAAAGATGACATTAAGATTGAAGGAC	480
Qy	1106	GAATTTGCAACAGAAAGATTCAACGAGTGGAAAAATTTGATTACATTCACAT	1165
Db	481	GTGTTCTGCAATGAAAGAAATTAAGTTA--AAAATCTTGTTTATCAATCAATCAGG	537
Qy	1166	TTGCAAGATTGTTCCAAAAGAAACAAGAGTGCACCAATTAATAGTTTAAAGCACTAGATA	1225
Db	538	AACTAACTGGGTTTAAGGAAGTAACGAGAGCCCTTCCATCTTGACTATCCCATCAATT	597
Qy	1226	ACTGGCTGTGATCTTAACGGGAGAAATTATGAAGCAGCAATTCATGATACAAATTC	1285
Db	598	TCTGGGCACTTTTTCACCAAGAGACATGACAGGCTCGAGAACTGCTCAACATG	657
Qy	1286	TATTTAAAGTTACACAGCCATGGGCATGCAATGAGAAAGCAATATGATTTGAAGT--	1343
Db	658	TGAGAAATATAGCGGCCCATTTGGCATGCTATGAGACCCACCGGCTGGGTTGAACATA	717
Qy	1344	-GGATGACAGAACTGAAAGCTTAATAGAGCTTAACAGCAAAAGTCAAGCAG-----	1396
Db	718	AGGATGACCGAATAAGACTTAATGTCAGAACCAATTCATTCACGTTAGAGCTGAGGGGA	777
Qy	1397	ACACCGAATAGTGTCTGTCTGTTGTCAAGTAATCGAAGACAAATACAGATCTATTA	1456
Db	778	AATATCAAGATGTTGTTTTCATCAATCATAGGGCCACGATGAGATCTATAGGGCCATCA	837
Qy	1457	AAAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGTGTGGGCCGAACCTTAG	1516
Db	838	AGAGAGTGTGCTGTGTGACGTCCCAAGTCCCTCCAGATGTTCATGTTCGAACATTC	897
Qy	1517	GCAAAACAGAAACTGTGATGAGCCATTTGCTACAAAGATTGCCCTACAGATGAATTCAGAA	1576
Db	898	GTCAGCCCAACAGGCTTCGGAATGTGGCCAGAAATTTTACTTCAGATTAACTGTAAT	957
Qy	1577	TGGAGAGAGACTCTGAGAGGTGTGACATCCCTCTGAATCTGTGATGATTCGTTGAGCATC	1636
Db	958	TGGTGTGTAGGCTCTGAGGAGTGGAAATTCCTCTAATAAGTTAATGTTGATCGGATGG	1017
Qy	1637	ATTGTACCATGATCATGACAGCTGGGCGAGGTCAATGCAAGATTTGTTGCCACATCA	1696
Db	1018	ATGTTTACATGACCCCAATGAAGGATCGCTCGGTGGTTGGCTTCGTGGCAAGCATCA	1077
Qy	1697	ATGAAGGATGACCCCGTGTCTCAAGCTGCATATTTAAGATTAAGGACAGAGAGCTGG	1756
Db	1078	ATCTCAACCTCACAAAATGTTATTTCCGGGTGTTGTCAGATGCCGATCAGAGATTCG	1137
Qy	1757	TAAATGGGCTCAAGTCTGCCTCAAGCGGCTTGAAGGCTTGAATAGCTCAATGAGT	1816

Db	1138	TGACAGCCTGAAAGCTATGCTCGTGGCTCCTTAAAAAAAGTTTATGAGGTGAACCACT	1197
Qy	1817	ACATGCCAGCCGGATCATCGTGTATCCGCAGATGCGGTAGAGACGGCCAGCTGAAAAACAC	1877
Db	1198	GTCTACCAAGAAAGATTGGTGTGTACCGTGTATGAGAGTGTCTGATGGCCAACTGAAAGACAG	1257
Qy	1877	TGCTGAATCTAGAAAGTGGCCACAGTTTGTGATGCTTAAATCCATTGATGAGGTATCA	1938
Db	1258	TTGCCACACTTATGAGATCTCTCAACTACAGAAAGTGTGTTGAAAGCTTTTGAAG---ATTATTC	1311
Qy	1937	ACCCTAGACTAACGGTAAATTGAGTGAAGAAAGAGTGAACAACAGATTTTGTGCTCAGT	1399
Db	1315	AGCCCAAGATGGTGGTGTGTTGATGATTCAAGAAAGATATGATCTATATCTGCGCTG	1377
Qy	1997	CTGAGAGAAAGACTTCAGAAATCCACTCTCTGAAACAGTTATGATGATGAGGTATCCAGAC	2058
Db	1375	CTCTCAGAACTTTGTATCTCCACTCTCGAAGCTGTGTATCATACAAATTAACAAGCT	1433
Qy	2057	CAGAAAGTATGACTTTTATTCGTAGACCAAGCTGTGAAAGTGTAGTGTCTTCCCA	2111
Db	1435	GTGATGGGTGGAATTTCTATCTTCTGGCCATCATATGACGAGGCGTGTGCATCTCTA	1497
Qy	2117	CACATTAACATGTATCTATATGACACACAGGGCGCTAAGCCACACATACAGCGCTTGA	2177
Db	1495	CCCATATATCTGTGTCTTCACACACCGCAACCTGAGCCCTGATCATATGACAGAGCTGA	1555
Qy	2177	CCTACAGAGTGTGCCCATCTATTAACAAGTGGCCAGAGTGTCAATGGTGTCTCTGCTCTT	2233
Db	1555	CTTTCAACTGTGACCAATGTAAGTGGCTGTGACCAATCAGAGTTCAGAGCTTCTT	1611
Qy	2237	GCCAGTACGCCCAACAGAGTGGCTTTCTTGTGGCCAGAGTATTCACAGAGGCAATC	2297
Db	1615	GCAAGTATGCCCAAGAGTGAAGTTCCTGTCAAGACACATCTTGATCATGACACAGCA	1677
Qy	2297	TGTCATCTGCAACCGCCTTTACTACCTCTAA	2328
Db	1675	TCCAGCTGTGCGAAGACCTGTCTTCTCTGTGA	1706
RESULT 15			
AB110571			
ID	AB110571	standard; cDNA; 2838 bp.	
XX	AC	AB110571;	
XX	DT	26-MAR-2002 (first entry)	
DB	XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.	
XX	XX	Drosophila; developmental biology; cell signalling; insecticide;	
XX	XX	pharmaceutical; gene; ss.	
XX	OS	Drosophila melanogaster.	
XX	PN	MO200117042-A2.	
XX	PD	27-SEP-2001.	
XX	PF	23-MAR-2001; 2001WO-US009231.	
XX	PR	23-MAR-2000; 2000US-0191637P.	
XX	PR	11-JUL-2000; 2000US-00614150.	
XX	PA	(PEKE ) PE CORP NY.	
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX	DR	WPI; 2001-656860/75.	
XX	DR	P-PSDB; ABB6468.	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.		



Wed May 12 09:50:20 2004

us-10-043-774b-1.rng

Page 20

```
Db 2277 TAAACAGCGAGTGAAACCCCTTAAGGACAGGCTCGACGAAATTTACAAATCAGCTGCA 2336
QY 1940 CTAGACTPAACGTAATTGTGTGTAAGAAAAGAGTGAAACACAGATTTTTGTGCTGACTG 1999
Db 2337 AACAGAGGGGCTGTGCGATACATTTATATTTGTATCCAGCGCATTAATTCGCTACT 2396
QY 2000 GAGGAAGACTTCAGATCCACTTCCTGGAACAGTTATGTATGATGAGGTTACGAGACCAG 2059
Db 2397 TTAAGGAGCATCGCAACCCAGTTCGGGCACTGTAGTCGATGACGTTATTAACCTTGCCAG 2456
QY 2060 AATGATGACTTTTATCGTGAGCCAGGCTGTGAGAGTGTAGTGTTCCTCCACAC 2119
Db 2457 AGCGCTAGACTTCTTCCTAGTGTGCCAGGCTGTGAGATAGGACTGTGTGCTGCTACCA 2516
QY 2120 ATTACATGTCTATGACCAACGCGGCTGAAAGCCAGACCATACAGCGCTGACCT 2179
Db 2517 GCTACAAATGTATTTCTGACCAACATGGGACTAAAGCCGATTAAGCTGACAGATGCTCTCT 2576
QY 2180 ACAAGCTGTGCAATCATATACACTGCGCAGGTGTCAATTCGTCTCTGCTCTGCGC 2239
Db 2577 ATTAAGATGACCATATGTACTACAAATTACAGCGGACCATACAGGTCCCGCTGTGCGC 2636
QY 2240 AGTAAGCCGACAGCTGCGCTTTCTTGTGGCCAGAGTATTACAGAGACCCAAATCTGT 2299
Db 2637 ATTAGGCCCAAAATGGCTTCTCTGTAAGCCGAAATCATTAATCGTGGCCTTCAGCAG 2696
QY 2300 CACTGTCAAACCGCCTTAACGCTTA 2328
Db 2697 GACTGCAGAAATCAATGTACTTTTGTAA 2725
```

Search completed: May 8, 2004, 11:37:40  
Job time : 936 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 09:23:26 ; Search time 9065 Seconds

(without alignments)  
1131.008 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328  
Sequence: 1 atgaccttcgttgtaaac.....accgccttactactctaa 2328

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Genbml:\*

1: gb\_ba:\*

2: gb\_hlg:\*

3: gb\_in:\*

4: gb\_ov:\*

5: gb\_ov:\*

6: gb\_ov:\*

7: gb\_ov:\*

8: gb\_ov:\*

9: gb\_ov:\*

10: gb\_ov:\*

11: gb\_ov:\*

12: gb\_ov:\*

13: gb\_ov:\*

14: gb\_ov:\*

15: gb\_ov:\*

16: gb\_ov:\*

17: gb\_ov:\*

18: gb\_ov:\*

19: gb\_ov:\*

20: gb\_ov:\*

21: gb\_ov:\*

22: gb\_ov:\*

23: gb\_ov:\*

24: gb\_ov:\*

25: gb\_ov:\*

26: gb\_ov:\*

27: gb\_ov:\*

28: gb\_ov:\*

29: gb\_ov:\*

30: gb\_ov:\*

31: gb\_ov:\*

32: gb\_ov:\*

33: gb\_ov:\*

34: gb\_ov:\*

35: gb\_ov:\*

36: gb\_ov:\*

37: gb\_ov:\*

38: gb\_ov:\*

39: gb\_ov:\*

40: gb\_ov:\*

41: gb\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2328	100.0	2328	9	AF264004	AF264004 Homo sapi
2	2316.2	99.5	3421	9	AF387507	AF387507 Homo sapi
3	2313	99.4	3362	9	AF104260	AF104260 Homo sapi
4	2313	99.4	3394	9	BC028581	BC028581 Homo sapi
5	2200.2	94.5	2579	6	AX747944	AX747944 Sequence
6	2200.2	94.5	2579	6	AX747944	AX747944 Sequence
7	1684.6	72.4	3910	10	AB033604	AB033604 Homo sapi
8	1684.6	72.4	4064	10	AF438405	AF438405 Mus muscu
9	1013.2	43.5	3273	5	AF36369	AF36369 Dario rer
10	913.2	39.2	3504	9	AB079368	AB079368 Homo sapi
11	680	29.2	3006	9	BC031060	BC031060 Homo sapi
12	680	29.2	3138	6	AX748151	AX748151 Sequence
13	680	29.2	3138	9	AK093475	AK093475 Homo sapi
14	616.8	26.5	2456	6	AX834401	AX834401 Sequence
15	616.8	26.5	2456	9	AK096837	AK096837 Homo sapi
16	568.6	24.4	3057	3	AY014899	AY014899 Strongylo
17	503.2	21.6	3458	3	AK116906	AK116906 Ciona int
18	499.4	21.5	2605	3	AY014800	AY014800 Strongylo
19	460.6	19.8	4913	10	AB032605	AB032605 Mus muscu
20	439.8	18.9	3545	9	AB079367	AB079367 Homo sapi
21	439.8	18.9	3610	9	BC025995	BC025995 Homo sapi
22	435.2	18.7	3620	6	AX714218	AX714218 Sequence
23	435.2	18.7	3620	6	AX714218	AX714218 Sequence
24	435.2	18.7	3620	9	AK056418	AK056418 Homo sapi
25	381	16.4	3649	6	AX247684	AX247684 Sequence
26	381	16.4	3659	10	AF285586	AF285586 Mus muscu
27	379.4	16.3	1723	3	AY014901	AY014901 Strongylo
28	330.2	14.2	1455	5	BC049495	BC049495 Dario rer
29	330	14.2	2272	6	AX879653	AX879653 Sequence
30	330	14.2	2272	6	BD157952	BD157952 Primer fo
31	330	14.2	2272	9	AK027497	AK027497 Homo sapi
32	310.6	13.3	2761	3	AF145680	AF145680 Drosophi
33	286.4	12.3	300	6	BD212731	BD212731 Novel hum
34	261.8	11.2	2982	3	BT011097	BT011097 Drosophi
35	257	11.0	3047	3	AF104354	AF104354 Drosophi
36	255.4	11.0	3054	3	BT011138	BT011138 Drosophi
37	237.2	10.2	1733	6	AX876238	AX876238 Sequence
38	237.2	10.2	1733	6	BD156024	BD156024 Primer fo
39	237.2	10.2	1733	9	AK001213	AK001213 Homo sapi
40	222.4	9.6	63578	9	AC127071	AC127071 Homo sapi
41	220.8	9.5	193697	2	AC025837	AC025837 Homo sapi
42	203.4	8.7	61995	2	AC090565	AC090565 Homo sapi
43	184	7.9	7781	3	DMDNARPL9	X94613 D.melanog
44	184	7.9	37980	3	AC005116	AC005116 Drosophi
45	184	7.9	78507	2	AC019704	AC019704 Drosophi

# ALIGNMENTS

RESULT 1

AF264004 AF264004 2328 bp. mRNA linear PRI 30-APR-2003

LOCUS Homo sapiens HIWI mRNA, complete cds.

DEFINITION AF264004

ACCESSION AF264004

VERSION AF264004.1 GI:15216446

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2328) Shatma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N., Weller,K.P. and Hoffman,R.

AUTHORS Human CD34+ stem cells express the hiwi gene, a human homologue of

Pred. No. is the number of results predicted by chance to have a

JOURNAL the Drosophila gene piwi  
 MEDLINE Blood 97 (2), 426-434 (2001)  
 20584579  
 PUBMED 11154219  
 REFERENCE 2 (bases 1 to 2328)  
 AUTHORS Sharpe, A.K. and Hoffman, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,  
 Chicago, IL 60607, USA

FEATURES  
 Source Location/Qualifiers

1..2328  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cbase\_type="cbaseis"  
 1..2328  
 /codon\_start=1  
 /product="piwi"  
 /protein\_id="AAK92281.1"  
 /translation="MTIGVNTROLNIDHYKESKSGGIVLSTNHFRLSRPOMALY  
 /db\_xref="GI:15216447"

CDS

1  
 ATGATCTTGTGTTGTAACACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGGT 60  
 1 ATGATCTTGTGTTGTAACACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGGT 60  
 61 TCTTACGAGCATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 120  
 61 TCTTACGAGCATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 120  
 121 TGGGCTTATATCAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTAATGAC 180  
 121 TGGGCTTATATCAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTAATGAC 180  
 181 TCGAGCTCTTCTTTCAACAGAGATCTAATGAGAAAGTCAATGCTTTTATGAGAAAG 240  
 181 TCGAGCTCTTCTTTCAACAGAGATCTAATGAGAAAGTCAATGCTTTTATGAGAAAG 240  
 181 TCGAGCTCTTCTTTCAACAGAGATCTAATGAGAAAGTCAATGCTTTTATGAGAAAG 240  
 241 ATATATTTTATCTTAAAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 300  
 241 ATATATTTTATCTTAAAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 300  
 301 AATGAGAGAGATGAGAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 360  
 301 AATGAGAGAGATGAGAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 360  
 361 ACTTGTGTCAGTTCTAATATTTTATGAGAGCTTTGAGAAATCAATGATTTGCA 420  
 361 ACTTGTGTCAGTTCTAATATTTTATGAGAGCTTTGAGAAATCAATGATTTGCA 420  
 421 CAATGAGAGAGATTTATTAACCAATGACCTAATGATTTGCAATGACCTAATGACCTA 480  
 421 CAATGAGAGAGATTTATTAACCAATGACCTAATGATTTGCAATGACCTAATGACCTA 480  
 481 GTGATTTGGCTGCTCACTCTTCACTGATGAGAAACAGCATCATGCTCTGC 540

ORIGIN

Query Match 100.0%; Score 2328; DB 9; Length 2328;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATCTTGTGTTGTAACACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGGT 60  
 DB 1 ATGATCTTGTGTTGTAACACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGGT 60  
 QY 61 TCTTACGAGCATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 120  
 DB 61 TCTTACGAGCATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 120  
 QY 121 TGGGCTTATATCAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTAATGAC 180  
 DB 121 TGGGCTTATATCAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTAATGAC 180  
 QY 181 TCGAGCTCTTCTTTCAACAGAGATCTAATGAGAAAGTCAATGCTTTTATGAGAAAG 240  
 DB 181 TCGAGCTCTTCTTTCAACAGAGATCTAATGAGAAAGTCAATGCTTTTATGAGAAAG 240  
 QY 241 ATATATTTTATCTTAAAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 300  
 DB 241 ATATATTTTATCTTAAAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 300  
 QY 301 AATGAGAGAGATGAGAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 360  
 DB 301 AATGAGAGAGATGAGAGATCACTAATGACCTAATGACCTAATGACCTAATGACCTA 360  
 QY 361 ACTTGTGTCAGTTCTAATATTTTATGAGAGCTTTGAGAAATCAATGATTTGCA 420  
 DB 361 ACTTGTGTCAGTTCTAATATTTTATGAGAGCTTTGAGAAATCAATGATTTGCA 420  
 QY 421 CAATGAGAGAGATTTATTAACCAATGACCTAATGATTTGCAATGACCTAATGACCTA 480  
 DB 421 CAATGAGAGAGATTTATTAACCAATGACCTAATGATTTGCAATGACCTAATGACCTA 480  
 QY 481 GTGATTTGGCTGCTCACTCTTCACTGATGAGAAACAGCATCATGCTCTGC 540

DB 481 GTGATTTGGCTGCTCACTCTTCACTGATGAGAAACAGCATCATGCTCTGC 540  
 QY 541 ACTGAGCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 600  
 DB 541 ACTGAGCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 600  
 QY 601 TATCATGAG 660  
 DB 601 TATCATGAG 660  
 QY 661 GTTGTCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 720  
 DB 661 GTTGTCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 720  
 QY 721 AATCCCAAG 780  
 DB 721 AATCCCAAG 780  
 QY 781 AGGAG 840  
 DB 781 AGGAG 840  
 QY 841 AAG 900  
 DB 841 AAG 900  
 QY 901 TGTATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 960  
 DB 901 TGTATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 960  
 QY 961 GCGCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1020  
 DB 961 GCGCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1020  
 QY 1021 TACATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1080  
 DB 1021 TACATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1080  
 QY 1081 TCCATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1140  
 DB 1081 TCCATCTTACGAGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1140  
 QY 1141 AAAACATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1200  
 DB 1141 AAAACATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1200  
 QY 1201 TTAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1260  
 DB 1201 TTAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1260  
 QY 1261 GCGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1320  
 DB 1261 GCGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1320  
 QY 1321 AGAAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1380  
 DB 1321 AGAAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1380  
 QY 1381 CAAGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1440  
 DB 1381 CAAGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1440  
 QY 1441 AATAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1500  
 DB 1441 AATAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1500  
 QY 1501 GTGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1560  
 DB 1501 GTGAGCAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1560  
 QY 1561 CAGATGAATTTGATTAAGTCTTGAAGTGAAGTCTTTGAGATTTCAATGCTTCACTT 1620



Db	1561	CAGATGAAC	TGCAGAGTGGAGGAGAGGTCTGGAGGGGTGACATCCCTGAACTGCTGTG	1520
Qy	1561	ATGATCGTTGGCAT	TGATTTGATCCATGACATGACAGCTGGGCGGAGGTCAATGCGAGGA	1580
Db	1621	ATGATCGTTGGCAT	TGATTTGATCCATGACATGACAGCTGGGCGGAGGTCAATGCGAGGA	1580
Qy	1681	TTTGTGGCAGCAT	CAATGAAGAGGTGACCGCGTGGTTCTCAGCGCTGCATTTTCAGGAT	1740
Db	1681	TTTGTGGCAGCAT	CAATGAAGAGGTGACCGCGTGGTTCTCAGCGCTGCATTTTCAGGAT	1740
Qy	1741	AGAGGACAGGAG	CTGTGATGAGGCTCAAACTCTGCTGCAAGCGCTCTTAAGGGCTTTGG	1800
Db	1741	AGAGGACAGGAG	CTGTGATGAGGCTCAAACTCTGCTGCAAGCGCTCTTAAGGGCTTTGG	1800
Qy	1801	AATAGCTGCAT	TGATGATCATGCGCCAGCGGATCATCGTACCGGATGGCGTAGAGAC	1860
Db	1801	AATAGCTGCAT	TGATGATCATGCGCCAGCGGATCATCGTACCGGATGGCGTAGAGAC	1860
Qy	1861	GGCCAGCTGAAA	CACTGTGTAACCTGTGAACCTGCAAGATGCGCAGTTTTGGATTGTCTTAAATCC	1920
Db	1861	GGCCAGCTGAAA	CACTGTGTAACCTGTGAACCTGCAAGATGCGCAGTTTTGGATTGTCTTAAATCC	1920
Qy	1921	ATTGTGAAGGT	TACACCTTGAATACCGTTAATGTGTGAGAAAGAGTAAAGACC	1980
Db	1921	ATTGTGAAGGT	TACACCTTGAATACCGTTAATGTGTGAGAAAGAGTAAAGACC	1980
Qy	1981	AGATTTTTTGGCT	CAGCTCGAGGAAAGCTTGAGATCCACTCTCGAACAGATTATGAT	2040
Db	1981	AGATTTTTTGGCT	CAGCTCGAGGAAAGCTTGAGATCCACTCTCGAACAGATTATGAT	2040
Qy	2041	GTAGAAGTTAC	CAAGCCAGAAATGTATGACTTTTTTAATCGTAGCGCAGAGCTGTGAAAGT	2100
Db	2041	GTAGAAGTTAC	CAAGCCAGAAATGTATGACTTTTTTAATCGTAGCGCAGAGCTGTGAAAGT	2100
Qy	2101	GGTAGTGTTC	CTCCCAACATTAACAATGATGACAAACAGCGGCGTGAAGCCAGAC	2160
Db	2101	GGTAGTGTTC	CTCCCAACATTAACAATGATGACAAACAGCGGCGTGAAGCCAGAC	2160
Qy	2161	CACATACAGC	GGTTGACCTTACAGCTGTGCAATCTAATTAACAATGCGCCAGGTGCAAT	2220
Db	2161	CACATACAGC	GGTTGACCTTACAGCTGTGCAATCTAATTAACAATGCGCCAGGTGCAAT	2220
Qy	2221	CGTGTTCCTGCT	CTCCCTTGGCAGAGCGCCACAAGCTGGCTTTCTGTGTGGCAGAGTAT	2280
Db	2221	CGTGTTCCTGCT	CTCCCTTGGCAGAGCGCCACAAGCTGGCTTTCTGTGTGGCAGAGTAT	2280
Qy	2281	CACAGAGAGCC	CAAAATCTGCTCACTGTCAAAACCGGCTTTACTACTCTTAA	2328
Db	2281	CACAGAGAGCC	CAAAATCTGCTCACTGTCAAAACCGGCTTTACTACTCTTAA	2328
RESULT 2				
AF387507				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				

[illegible]







SKTSGSGLIVLSTNHPRLTSRPMOALYOHYDINPLMEARLSALLFOHEDLIGK  
HAFQTLIFLPRKLOQKVEKSRKNGEDRITITLNLNPEPSPCLPFYNIIFER  
LTKMNDQIGRNTNPNYDIDIPSHLVITPFTTSLIOWENIMLCTDYSHVLRS  
ETVLDPMFNHOTEERKPOVSKEILGLVLTINKTKTRVDIMDMONPSKTEPK  
ADGSEVSELEYRYKQXNOEITDLKQPVLSQPKRRGGITLPEPALIPBLCTTLU  
TDKRMNDVWKMDLAVHTRLEPQORQEVGLIDYIKNDVQSELWDWGLSEPSNL  
SFSGRILLOTETKIHOGKTFDYNPQADWSKETRAPILSVPLDMLLIYTRRYEA  
NSLIONLPTKIPAMGOMKKAIMEVDRETAURLVLOQKTAADTOIVCLLSNRD  
KYDAIKXPCHDCPTSPQCVARLTCGQTMATATKIALCMCKMGEMLRVIDIPK  
LMTVIGDCHMDTGRSISAGFASISIEGTWRPSCITPDROGELVDGLKVLQAA  
LRAMVSCHEPMSRLIVRVDGSDQLTVNREYRPOGLDKSISGGVYPLVIVY  
KRVNTRFPAOSGRLONPPEYVIVDVTPEYDFIVASQARSVSSEPHNIVY  
DNLGKPHIQRILTYKLCHIIYMWGVIRVAPCQYAKRLAFLVQSITHRPNLSIN  
RLYYL"

misc\_feature  
900..1310  
/note="PAZ; Region: PAZ domain. This domain is named PAZ  
alter the proteins Pw1 Argonaut and Zwille. This domain  
is found in two families of proteins that are involved in  
post-transcriptional gene silencing. These are the Pw1  
family and the Dicer family, that includes the Carpel  
factory protein. The function of the domains is unknown  
but has been suggested to mediate complex formation  
between proteins of the Pw1 and Dicer families by  
hetero-dimerisation"  
/db\_xref="CDD:Pfam02170"  
1728..2609  
/note="Pw1; Region: Pw1 domain. This domain is found in  
the protein Pw1 and its relatives. The function of this  
domain is unknown"  
/db\_xref="CDD:Pfam02171"

ORIGIN  
Query Match 99.4%; Score 2313; DB 9; Length 3394;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

8 TTGGTGTGAACACAGGCAAGTCTAGACATGTTAAGATCAAAAAGAGTCTTCAG 67  
334 TTGGTGTGAATAAAGGCAAGCTAGACATGTTAAGATCAAAAAGAGTCTTCAG 393  
68 GCATTAATGAAGTAAAGCACTAACCATTTCCGGGTGAACATCCGCTCCCAAGTGGCCT 127  
394 GCATTAATGAAGTAAAGCACTAACCATTTCCGGGTGAACATCCGCTCCCAAGTGGCCT 453  
128 TATATCAGTATCACTGATTAACCCACTGATGGAAGGCAAGAACTCCGTTGAGTGC 187  
454 TATATCAGTATCACTGATTAACCCACTGATGGAAGGCAAGAACTCCGTTGAGTGC 513  
188 TTCTTTTCAACAGAGATCTAATGGAAGTGTCAATGCTTTTGAAGAGATTAAT 247  
514 TTCTTTTCAACAGAGATCTAATGGAAGTGTCAATGCTTTTGAAGAGATTAAT 573  
248 TTTTACCTAAAGCTACAGCAAAAGTTACTGAAGTCTTTTGAAGAGATTAAT 307  
574 TTTTACCTAAAGCTACAGCAAAAGTTACTGAAGTCTTTTGAAGAGATTAAT 633  
308 AGAGTGAAGATAGATCACTTAACAATGAAGTCCACCTACATCAAACTTGT 367  
634 AGAGTGAAGATAGATCACTTAACAATGAAGTCCACCTACATCAAACTTGT 693  
368 TGCAGTCTAATATTAATTTTCAAGAGGCTTTGAAAATCAGAAATTTGCAAAATG 427  
694 TGCAGTCTAATATTAATTTTCAAGAGGCTTTGAAAATCAGAAATTTGCAAAATG 753  
428 GACGAATTAATTAACCCAAATGACCAATGATTTTCAAGTCAAGGTGGTGAAT 487  
754 GACGAATTAATTAACCCAAATGACCAATGATTTTCAAGTCAAGGTGGTGAAT 813  
488 GAGCGTGTCTACTACTTCACTTCAAGTGAAGAAACAGATCATGCTGTGCACTAG 547  
814 GAGCGTGTCTACTACTTCACTTCAAGTGAAGAAACAGATCATGCTGTGCACTAG 873  
548 TTAGCATTAAGTCTTGAAGTGAAGTCTTTTGAATTCATGTTCAACTTATCATC 607

874 TTAGCATTAAGTCTTGAAGTGAAGTCTTTTGAATTCATGTTCAACTTATCATC 933  
608 AGACAGAGAACTAATATTTCAAGAAACAAGTTTCCAAAGAACTAATAGTTTAGTTGTC 667  
934 AGACAGAGAACTAATATTTCAAGAAACAAGTTTCCAAAGAACTAATAGTTTAGTTGTC 993  
668 TTACCAAGTATTAACAAATTAAGCACTAAGAGTGAATTAATTTGACTGGAGCAAGATCCA 727  
994 TTACCAAGTATTAACAAATTAAGCACTAAGAGTGAATTAATTTGACTGGAGCAAGATCCA 1053  
728 AGAGCACTTTAAGAAAGCCGACGAGCTGAAGTCAAGTCTTTTGAATTAACAAAGAGC 787  
1054 AGAGCACTTTAAGAAAGCCGACGAGCTGAAGTCAAGTCTTTTGAATTAACAAAGAGC 1113  
788 AATPACAACAAAGATACCGACTTGAAGAGCTCTTGTGACCGCCCAAGAGAA 847  
1114 AATPACAACAAAGATACCGACTTGAAGAGCTCTTGTGACCGCCCAAGAGAA 1173  
848 GGGGGGGCCCTGGGGGGGACCTGCGAGGGCTCCATGCTCATTCCTGAGCTGTCTATC 907  
1174 GGGGGGGCCCTGGGGGGGACCTGCGAGGGCTCCATGCTCATTCCTGAGCTGTCTATC 1233  
908 TTACAGTCTTAAGTAAATGCGTATGATTTTAACTGATGAAGAACTTAGCCGCTTC 967  
1234 TTACAGTCTTAAGTAAATGCGTATGATTTTAACTGATGAAGAACTTAGCCGCTTC 1293  
968 ATPACAAGCTTAATCCGAGCAAGAGGAGTGAAGGAGGAGCACTGATGATTAATTC 1027  
1294 ATPACAAGCTTAATCCGAGCAAGAGGAGTGAAGGAGGAGCACTGATGATTAATTC 1353  
1028 AATAAAAGATTAATGTTCAAAAGGAGCTTGAAGCTGGGTTGAGCTTTTCAACT 1087  
1354 AATAAAAGATTAATGTTCAAAAGGAGCTTGAAGCTGGGTTGAGCTTTTCAACT 1413  
1088 TACTGCTCTTCAAGAAAGATTTTGAAGAGAAAGATTCACCAAGTGAAGAAACAT 1147  
1414 TACTGCTCTTCAAGAAAGATTTTGAAGAGAAAGATTCACCAAGTGAAGAAACAT 1473  
1148 TTGATTAATCAATCAAAATTTGCAAGTGTGTCAAAGAAACAAGGTTGACATTAATTA 1207  
1474 TTGATTAATCAATCAAAATTTGCAAGTGTGTCAAAGAAACAAGGTTGACATTAATTA 1533  
1208 GTGTTAAGCCATGATTAATCTGCTGTTGATCTTAAGCCGAGAAATTAAGAGAGCA 1267  
1534 GTGTTAAGCCATGATTAATCTGCTGTTGATCTTAAGCCGAGAAATTAAGAGAGCA 1593  
1268 ATTCAATTGATCAAAATCTATTTAAAGTTTACACAGCCATGGGCAATGCAAAATG 1327  
1594 ATTCAATTGATCAAAATCTATTTAAAGTTTACACAGCCATGGGCAATGCAAAATG 1653  
1328 CAATTAATTAATGAATGATGAAGAACTGAAGCTTCTTAAGAGTCTTAAGCAAGAAAG 1387  
1654 CAATTAATTAATGAATGATGAAGAACTGAAGCTTCTTAAGAGTCTTAAGCAAGAAAG 1713  
1388 TCAAGCAAGACCAACGATGATGCTGCTGCTGTTGTCAGATTAAGCAAGAAATG 1447  
1714 TCAAGCAAGACCAACGATGATGCTGCTGCTGTTGTCAGATTAAGCAAGAAATG 1773  
1448 ATGCTATTAAATAATACCTGTGTGACAGATGTCCTTAAGCTCAAGTCAAGTGTGGGCC 1507  
1774 ATGCTATTAAATAATACCTGTGTGACAGATGTCCTTAAGCTCAAGTCAAGTGTGGGCC 1833  
1508 GAACCTTAAGCAAAACAGAACTGTCAGAGCCATGCTTACAAAGATTTGCCATAGATGA 1567  
1834 GAACCTTAAGCAAAACAGAACTGTCAGAGCCATGCTTACAAAGATTTGCCATAGATGA 1893  
1568 ACTGCAAGATGGAAGAGAGCTCTGAGAGGTGAGCACTCCCTGAAAGCTGTGATGATG 1627  
1894 ACTGCAAGATGGAAGAGAGCTCTGAGAGGTGAGCACTCCCTGAAAGCTGTGATGATG 1953  
1628 TTGGCATGATTTGTTTCAATGATGACAGCTGGGGGAGAGTCAATCGCAGATTTGTTG 1687



Db 1954 TTGGCATTCATTTGTTACATGACATGACGCTGGGCGGAGCTCAATCGCAGATTTGTTG 2013  
 Qy 1688 CCAGCATCAATGAGAGGATGACCCGCTGGTCTCGCTGCATTTTTCAGATTAAGAGAC 1747  
 Db 2014 CCAGCATCAATGAGAGGATGACCCGCTGGTCTCGCTGCATTTTTCAGATTAAGAGAC 2073  
 Qy 1748 AGGAGCTGGTATAGGCTCAAGTCTGGCTGCAAGCGGCTTGAAGGCTTGAATAGCT 1807  
 Db 2074 AGGAGCTGGTATAGGCTCAAGTCTGGCTGCAAGCGGCTTGAAGGCTTGAATAGCT 2133  
 Qy 1808 GCAATAGTACATGACCCGAGATCATCTGTACCGGATGCGGTAGAGAGCGGCGAC 1867  
 Db 2134 GCAATAGTACATGACCCGAGATCATCTGTACCGGATGCGGTAGAGAGCGGCGAC 2193  
 Qy 1868 TGAATACATGCTGTAATCTAGAAAGTGGCCACATTTTGGATTGTCTAAATTCATTGGTA 1927  
 Db 2194 TGAATACATGCTGTAATCTAGAAAGTGGCCACATTTTGGATTGTCTAAATTCATTGGTA 2253  
 Qy 1928 GAGGTACACACCTAGACTAAGCTAATGTGTAGAGAGAGAGAGAGAGAGAGAGAGAG 1987  
 Db 2254 GAGGTACACACCTAGACTAAGCTAATGTGTAGAGAGAGAGAGAGAGAGAGAGAGAG 2313  
 Qy 1988 TTGCTCAGTCTGGAGAGAGACTTTCAGATTCCTCTGGAACAGTTATGATGTAGAG 2047  
 Db 2314 TTGCTCAGTCTGGAGAGAGACTTTCAGATTCCTCTGGAACAGTTATGATGTAGAG 2373  
 Qy 2048 TTACCAAGCAGAGATGCTATGATTTTATGCTGAGCAGGCTGTGAGAGATGCTATG 2107  
 Db 2374 TTACCAAGCAGAGATGCTATGATTTTATGCTGAGCAGGCTGTGAGAGATGCTATG 2433  
 Qy 2108 TTCTCCCAACATTAATGCTATGATGATGACAAAGCGGCTGGAAGCCAGCCATAC 2167  
 Db 2434 TTCTCCCAACATTAATGCTATGATGATGACAAAGCGGCTGGAAGCCAGCCATAC 2493  
 Qy 2168 AGGCTTGAACCTAGAGCTGTGACATGATTTTAACTGAGCGAGTGTGATGCTG 2227  
 Db 2494 AGGCTTGAACCTAGAGCTGTGACATGATTTTAACTGAGCGAGTGTGATGCTG 2553  
 Qy 2228 CTGCTCTTGCAGAGAGCGCCCAAGCTGCTTTTCTGTTGCGCAGAGTATTCACAG 2287  
 Db 2554 CTGCTCTTGCAGAGAGCGCCCAAGCTGCTTTTCTGTTGCGCAGAGTATTCACAG 2613  
 Qy 2288 AGCCAAATCTGTCACTGTCAAAACCGCTTTTACCTACTTAA 2328  
 Db 2614 AGCCAAATCTGTCACTGTCAAAACCGCTTTTACCTACTTAA 2654

RESULT 5  
 AX747944 2579 bp mRNA linear PAT 20-JUN-2003  
 LOCUS Sequence 1469 from Patent EPI308459.  
 AX747944  
 DEFINITION AX747944.1 GI:32132332  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS  
 1 Isegai, T., Sugiyama, T., Otsuki, T., Makamatsu, A., Sato, H., Ishii, S.,  
 Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,  
 Tamehika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and  
 Masuno, Y.

TITLE  
 JOURNAL  
 Patent: EP 1308459-A 1469 07-MAY-2003;  
 Helix Research Institute (JP); Research Association for  
 Biotechnology (JP)

FEATURES  
 SOURCE  
 1. 2579  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

ORIGIN

Query Match 94.5%; Score 2200.2; DB 6; Length 2579;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 8 TTGGTGAACACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGTTCTTCAG 67  
 Db 319 TTGGTGAATACAGGAGAACTTAGACATGTTAAAGATCAAAAACAGTTCTTCAG 378  
 Qy 68 GCATTAATAGTAAAGTTAAGACATTAACATTTCCGCTGACATCCGCTCCAGTGGGCT 127  
 Db 379 GCATTAATAGTAAAGTTAAGACATTAACATTTCCGCTGACATCCGCTCCAGTGGGCT 438  
 Qy 128 TATATCAGTATCACTATGATTAACCACTGATGAGAGCCAGAGACTCCGTCAGCTC 187  
 Db 439 TATATCAGTATCACTATGATTAACCACTGATGAGAGCCAGAGACTCCGTCAGCTC 498  
 Qy 188 TTCTTTTCAACAGAGATCTAATTTGAAAGTGTATGTTTGAAGAGAGATTAAT 247  
 Db 499 TTCTTTTCAACAGAGATCTAATTTGAAAGTGTATGTTTGAAGAGAGATTAAT 558  
 Qy 248 TTTTACCTAAAGACTACAGCAAAAGTTACTAGTGTATTTAGTAAGACCCGAGATGAG 307  
 Db 559 TTTTACCTAAAGACTACAGCAAAAGTTACTAGTGTATTTAGTAAGACCCGAGATGAG 618  
 Qy 308 AGGATGAGAGATPAACGATCACTTTAACAATGAACCTTCACTACATCAACCACTGTT 367  
 Db 619 AGGATGAGAGATPAACGATCACTTTAACAATGAACCTTCACTACATCAACCACTGTT 678  
 Qy 368 TGCAATCTAATTAATTAATTTTCAAGAGGCTTTTGAAGATCAAGAAATTTGCAAAATG 427  
 Db 679 TGCAATCTAATTAATTAATTTTCAAGAGGCTTTTGAAGATCAAGAAATTTGCAAAATG 728  
 Qy 428 GACGAAATTAATTAATCCCAATGATGATTTTCAAGTCAAGTGTGATTT 487  
 Db 739 GACGAAATTAATTAATCCCAATGATGATTTTCAAGTCAAGTGTGATTT 798  
 Qy 488 GGCCTGGCTTCACTACTTCATCTTCAATGATGATGATGATGATGATGATGATGATG 547  
 Db 799 GGCCTGGCTTCACTACTTCATCTTCAATGATGATGATGATGATGATGATGATGATG 858  
 Qy 548 TTACCAATTAATGCTTGAAGTGAAGCTGTTTGAATTTCAATGATGATGATGATG 607  
 Db 859 TTACCAATTAATGCTTGAAGTGAAGCTGTTTGAATTTCAATGATGATGATGATG 918  
 Qy 608 AGACAGAAAGACATTAATTTCAAGAAAGATTTCCAAAGAACTTAATGATTTAGTTTC 667  
 Db 919 AGACAGAAAGACATTAATTTCAAGAAAGATTTCCAAAGAACTTAATGATTTAGTTTC 978  
 Qy 668 TTACCAATTAATTAATTAAGATTAAGATGATGATGATGATGATGATGATGATGATG 727  
 Db 979 TTACCAATTAATTAATTAAGATTAAGATGATGATGATGATGATGATGATGATGATG 1038  
 Qy 728 AGAGCACTTTTAAGAAAGCGGAGCTCTAAGTCACTTCTTAAGTAACTTCAAGAGAG 787  
 Db 1039 AGAGCACTTTTAAGAAAGCGGAGCTCTAAGTCACTTCTTAAGTAACTTCAAGAGAG 1098  
 Qy 788 AATACCAACAGAGATCAAGCACTTGAAGAGCTGCTTGTGAGGAGCCAGCCAGAGAA 847  
 Db 1099 AATACCAACAGAGATCAAGCACTTGAAGAGCTGCTTGTGAGGAGCCAGCCAGAGAA 1158  
 Qy 848 GCGCGGCGCTTGGGAGAGCACTGCGAGGCGCTGCAATGCTCACTTCCAGAGCTGCTATC 907  
 Db 1159 GCGCGGCGCTTGGGAGAGCACTGCGAGGCGCTGCAATGCTCACTTCCAGAGCTGCTATC 1218  
 Qy 908 TTACAGTCTAATGATTAATTTGATGATTTTAACTGATGATGATGATGATGATGATG 967  
 Db 1219 TTACAGTCTAATGATTAATTTGATGATTTTAACTGATGATGATGATGATGATGATG 1278  
 Qy 968 ATACAGAGTAACTTCCAG 1027  
 Db 1279 ATACAGAGTAACTTCCAG 1338





## ORIGIN

DMSGKPDH1QRLTYKCHLYNMPVASTC"

Query Match 94.5%; Score 2200.2; DB 9; Length 2579;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACAAGGCGAAGCTTAAAGCAATGTAAGCAATCAAAAAGAGTTCTTCAAG 67  
 DB 319 TTGGTGTGAACAAGGCGAAGCTTAAAGCAATGTAAGCAATCAAAAAGAGTTCTTCAAG 378  
 QY 68 GCATTATAGTAGGTTAAGCACTAACCAATTCGCGCTGACATCCCGTCCCAAGTGGCCT 127  
 DB 379 GCATTATAGTAGGTTAAGCACTAACCAATTCGCGCTGACATCCCGTCCCAAGTGGCCT 438  
 QY 128 TATATCAATATCACTTGAATTAACCACTGATGGAAGCCAGAAAGCTCCGTTCAAGTC 187  
 DB 439 TATATCAATATCACTTGAATTAACCACTGATGGAAGCCAGAAAGCTCCGTTCAAGTC 498  
 QY 188 TTCTTTTCAACAGCAAGATCTAAATGGAAGTGTCACTTTTGTATGGAACGATATTAT 247  
 DB 499 TTCTTTTCAACAGCAAGATCTAAATGGAAGTGTCACTTTTGTATGGAACGATATTAT 558  
 QY 248 TTTTACCTTAAAGACTTACAGCAAAAAGTTACTGAAGTTTAAAGTAACCCGGAATGAG 307  
 DB 559 TTTTACCTTAAAGACTTACAGCAAAAAGTTACTGAAGTTTAAAGTAACCCGGAATGAG 618  
 QY 308 AGAGTGTGAAGTAACGATCACTTAAACAATGAACTTCACTACATCAACCACTTGT 367  
 DB 619 AGAGTGTGAAGTAACGATCACTTAAACAATGAACTTCACTACATCAACCACTTGT 678  
 QY 368 TGCAGTCTTAATAATAATTTTCAAGAGGCTTTGAAAATCAATGAATTTGCAAAAATTG 427  
 DB 679 TGCAGTCTTAATAATAATTTTCAAGAGGCTTTGAAAATCAATGAATTTGCAAAAATTG 738  
 QY 428 GACGAAATTAATAAACCAGAAATGACCAATTTGAATTTCAAGTCAAGTGGTGAATT 487  
 DB 739 GACGAAATTAATAAACCAGAAATGACCAATTTGAATTTCAAGTCAAGTGGTGAATT 798  
 QY 488 GGCCTGGCTTCACTCTTCATCTTCAGATGAAAACAGATCACTGCTGCACTGAG 547  
 DB 799 GGCCTGGCTTCACTCTTCATCTTCAGATGAAAACAGATCACTGCTGCACTGAG 858  
 QY 548 TTAGCCTTAATTCCTTGAAGTGAAGCTTTTGTGATTTCAATGTTCACTTATATC 607  
 DB 859 TTAGCCTTAATTCCTTGAAGTGAAGCTTTTGTGATTTCAATGTTCACTTATATC 918  
 QY 608 AACACAGAAAGATTAATTTCAAGAAACAATTTCCAAAAGTAATAGTTTATGTTTC 667  
 DB 919 AACACAGAAAGATTAATTTCAAGAAACAATTTCCAAAAGTAATAGTTTATGTTTC 978  
 QY 668 TTAACAAGTAATCAATTAAGCAATACAGGTGATGATTTGAACCTGGGACCAAGATCCCA 727  
 DB 979 TTAACAAGTAATCAATTAAGCAATACAGGTGATGATTTGAACCTGGGACCAAGATCCCA 1038  
 QY 728 AGAGCACTTTAAGAAAGCCGAGCGCTCTGAAGTCAAGTTCTTAAGAAATCAACAGAAAC 787  
 DB 1039 AGAGCACTTTAAGAAAGCCGAGCGCTCTGAAGTCAAGTTCTTAAGAAATCAACAGAAAC 1098  
 QY 788 AATACAAACCAAGATCAACGACTTGAAGCAGCTGTCTTGTCAAGCCAGCCCAAGAA 847  
 DB 1099 AATACAAACCAAGATCAACGACTTGAAGCAGCTGTCTTGTCAAGCCAGCCCAAGAA 1158  
 QY 848 GAGGGGGCCCTGGGGGGAACCTGACAGGGCTGCACTGCTCACTTCTGAGCTCTGATC 907  
 DB 1159 GAGGGGGCCCTGGGGGGAACCTGACAGGGCTGCACTGCTCACTTCTGAGCTCTGATC 1218  
 QY 908 TTAACAGTCTTAATTAATTAATGCTTAATTTTAACTGATGAAGACTTAAGCCGTTTC 967  
 DB 1219 TTAACAGTCTTAATTAATTAATGCTTAATTTTAACTGATGAAGACTTAAGCCGTTTC 1278  
 QY 968 ATAACAAGCTTAATTCAGAGCAAAAGGAGCGTGAATGGGACGACTCATTTGATTCATTC 1027

DB 1279 ATACAGAAGCTAATCCAGAGCAAAAGGCAAGCGTGAAGTGGACGACTCATTTGATTCATTC 1338  
 QY 1028 ATAAAAAGATTAATGTTCAAAAGGAGCTTCAAGACTGAGGCTTTGAGCTTTGATTCCAACT 1087  
 DB 1339 ATAAAAAGATTAATGTTCAAAAGGAGCTTCAAGACTGAGGCTTTGAGCTTTGATTCCAACT 1398  
 QY 1088 TACTGCTCTTCAAGAAAGATTTTGGCAACGAAAGATTTCAACCAAGTGGAAAAACAT 1147  
 DB 1399 TACTGCTCTTCAAGAAAGATTTTGGCAACGAAAGATTTCAACCAAGTGGAAAAACAT 1458  
 QY 1148 TTGATTAAGTCAACAAATTTGCAATGCTGCTCAAGAAACCAAGAGTGCACCATTAATTA 1207  
 DB 1459 TTGATTAAGTCAACAAATTTGCAATGCTGCTCAAGAAACCAAGAGTGCACCATTAATTA 1518  
 QY 1208 GTGTTAAGCCACTAGATTAATCTGCTGTTGATCTAATACCGAAAGAAATTTGAAAGACCA 1267  
 DB 1519 GTGTTAAGCCACTAGATTAATCTGCTGTTGATCTAATACCGAAAGAAATTTGAAAGACCA 1578  
 QY 1268 ATTCATGTAACAAATCTAATTTAAGTTACACCGCAATGGGCAATGCAAGAAAG 1327  
 DB 1579 ATTCATGTAACAAATCTAATTTAAGTTACACCGCAATGGGCAATGCAAGAAAG 1638  
 QY 1328 CAATATGATTAAGTGAATGATGAACAGAACTGAAGCTTAATTAAGTCTTACAGCAAAAG 1387  
 DB 1639 CAATATGATTAAGTGAATGATGAACAGAACTGAAGCTTAATTAAGTCTTACAGCAAAAG 1698  
 QY 1388 TCACAGCAACACCCAGATGTTGTTCTGTCTGTGTCAAGTAATGGAAAGCAAAATAGC 1447  
 DB 1699 TCACAGCAACACCCAGATGTTGTTCTGTCTGTGTCAAGTAATGGAAAGCAAAATAGC 1758  
 QY 1448 ATGCTATTAATAAAATCTGTTGATACAGATGTCCTTACCCCAAGTCAAGTGTGTGGGCC 1507  
 DB 1759 ATGCTATTAATAAAATCTGTTGATACAGATGTCCTTACCCCAAGTCAAGTGTGTGGGCC 1818  
 QY 1508 GAACTTTAGGCAAAACAGCAAACTGTCTATGCTATGCTCAAAAGATGCTCTACAGATGA 1567  
 DB 1819 GAACTTTAGGCAAAACAGCAAACTGTCTATGCTATGCTCAAAAGATGCTCTACAGATGA 1878  
 QY 1568 ACTGCAAGATGGGAGGAGAGCTCTGAGAGGTGGAACATCCCGTGAAGCTCGTATGATGC 1627  
 DB 1879 ACTGCAAGATGGGAGGAGAGCTCTGAGAGGTGGAACATCCCGTGAAGCTCGTATGATGC 1938  
 QY 1628 TTGGCATGATTTGTAACATGACATGACAGCTGGGGGAGGCTCAATCCGAGATTTGTTG 1687  
 DB 1939 TTGGCATGATTTGTAACATGACATGACAGCTGGGGGAGGCTCAATCCGAGATTTGTTG 1998  
 QY 1688 CCAGCATCAATGAAGGAGTGAACCGCTGCTCTCAAGCTGATTTTCAAGATGAAGAGAC 1747  
 DB 1999 CCAGCATCAATGAAGGAGTGAACCGCTGCTCTCAAGCTGATTTTCAAGATGAAGAGAC 2058  
 QY 1748 AGAGCTGTGAATGGGCTTCAAAAGTCTGCTGCAAGCGGCTCTGAAGGCTTGAATAGCT 1807  
 DB 2059 AGAGCTGTGAATGGGCTTCAAAAGTCTGCTGCAAGCGGCTCTGAAGGCTTGAATAGCT 2118  
 QY 1808 GCATGATGTAATGAGCCAGCGGATCATGCTGACCGGATGAGAGAGGCGACG 1867  
 DB 2119 GCATGATGTAATGAGCCAGCGGATCATGCTGACCGGATGAGAGAGGCGACG 2178  
 QY 1868 TGAAAACACTGATGAATCAAGAGTGCACAGTTTGTGATTTGATTTCTTAATTCATTTGTA 1927  
 DB 2179 TGAAAACACTGATGAATCAAGAGTGCACAGTTTGTGATTTGATTTCTTAATTCATTTGTA 2238  
 QY 1928 GAGGTTCAACCTTGAATCAACGCTTAATGTTGATGGAAGAAAGAGTGAACCAAGATTTT 1987  
 DB 2239 GAGGTTCAACCTTGAATCAACGCTTAATGTTGATGGAAGAAAGAGTGAACCAAGATTTT 2298  
 QY 1988 TTGCTCACTGTGAGGAAGACTTCAAGATCACTCTCTGGAACAGTTATTTGATGTAAGG 2047  
 DB 2299 TTGCTCACTGTGAGGAAGACTTCAAGATCACTCTCTGGAACAGTTATTTGATGTAAGG 2358  
 QY 2048 TTAACAAGCAAAATGTTGATGCTTTTATTCGAGACCGAGGCTGTGAAGAGTGTAGTG 2107  
 DB 2359 TTAACAAGCAAAATGTTGATGCTTTTATTCGAGACCGAGGCTGTGAAGAGTGTAGTG 2418





Db 517 TGGATCTATGTGAGAGCGACCAACCACTTCGGCTGACCTGGCCCAAGTGGGC 576  
Qy 126 CTATATCAATATCACTATGACTATTAACCACTGATGAGACCGAAGATCTCGTCAAC 185  
Db 577 CCTGATATGATACCATGACTACCAATCCCTGATGAGAGCCGGAAGCTTCCTCGC 636  
Qy 186 TCTTCTTTTCAACGAAAGATCTAATGGAAGTGCATGCTTTGATGAGAGATAT 245  
Db 637 ACTGCTCTTCCGACTGAAGACTCATTTGAAGGTCTATGCTTCAATGGGAATAT 696  
Qy 246 ATTTTACCTTAAGACTACAGCAAAAGTTACTGAAGTTTGAAGAACCCGAATGG 305  
Db 697 GTTTTACCTTAAGACTACAGCAAAAGTTACTGAAGTTTGAAGAACCCGAATGG 756  
Qy 306 AGAGATGTGAGAGATACGATCACTTAACAATGAATTCACCTCATATCAACAATGG 365  
Db 757 GGAACACGTGAGAGATACCTATCACTGACCAAGCTGCGCCCATCTGCCCCACTG 816  
Qy 366 TTGAGATCTAATTAATTAATTTGAGAGGCTTTGAAAATCAATGAATTTGCAAAAT 425  
Db 817 CCTGAGTCTAATATCATATCTTCAAGAGGCTTTGAAAATCAATGAATTTGCAAAAT 876  
Qy 426 TGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 485  
Db 877 TGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 936  
Qy 486 TTGGCTGGCTCACTATCTGATCTGATGATGAAAGACATCATGCTCTGCACTGA 545  
Db 937 CTGGCCGGCTTCACTATCTGATCTGATGATGAAAGACATCATGCTCTGCACTGA 996  
Qy 546 CGTTAGCCATAAAGTCTTGAAGTGAAGCTTTGATTTGATTTGATTTGATTTGAT 605  
Db 997 CGTCAAGCAAGAGTCTGCGAGAGAGCTGCTCAAGTCTCAATCTAATCA 1056  
Qy 606 TGAAGCAAGAACTAATTTCAAGAAAGTTTCAAGAAAGTTTCAAGAAAGTTTCAAG 665  
Db 1057 GAGAGCAAGAGCAAGTTTCAAGAAAGTTTCAAGAAAGTTTCAAGAAAGTTTCAAG 1116  
Qy 666 TCTTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 725  
Db 1117 TCTTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1176  
Qy 726 CAAGAGCACTTTAAGAAAGCGGCTCTGAATCAAGTCTTTAAGAAAGTTTCAAGAA 785  
Db 1177 AAGAGCACTTTAAGAAAGCGGCTCTGAATCAAGTCTTTAAGAAAGTTTCAAGAA 1236  
Qy 786 GGAATACAAAGAGATCAAGCACTTGAAGAGCTGCTGATCAAGCAAGCAAGCAAG 845  
Db 1237 GGAATACAAAGAGATCAAGCACTTGAAGAGCTGCTGATCAAGCAAGCAAGCAAG 1296  
Qy 846 AAGCGGGGCTTGGGGGCACTGCGCAGGCTTGCATCTCATTTCTGAGCTTGTGA 905  
Db 1297 GAGAGAGGCTTGGGGGCACTGCGCAGGCTTGCATCTCATTTCTGAGCTTGTGA 1356  
Qy 906 TCTTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 965  
Db 1357 TCTTCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1416  
Qy 966 TCAATACAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1025  
Db 1417 GGAACAGCGGCTGACCTGAGCAGCGGAGGCTGAGGCTGAGGCTGAGGCTGAGG 1476  
Qy 1026 TCAATACAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1085  
Db 1477 CCAAGAGATGAGATGAGAGAGCTTGAAGCTTGGGCTGAGGCTGAGGCTGAGG 1536  
Qy 1086 CTATGCTCTTCAAGAAAGTTTGAAGCAAGAAAGATTCAGAGGTTGAAGAAAC 1145  
Db 1537 CTGCTGCTCTTCAAGAAAGTTTGAAGCAAGAAAGATTCAGAGGTTGAAGAAAC 1596  
Qy 1146 ATTGATTAATCAATCAATTTGCAAGTTGCTCAAGAAAGCAAGGCTGCAATTAAT 1205  
Db 1597 GTTTGATTAACCAACCAATTTGCAAGTTGCTCAAGAAAGCAAGGCTGCAATTAAT 1656  
Qy 1206 TATGTTAAGCACTGATTAATGCTGTTGATCTAATGCGGAAGAAATTAATGAACAG 1265  
Db 1657 CAGCGTGAAGCACTGATTAATGCTGTTGATCTAATGCGGAAGAAATTAATGAACAG 1716  
Qy 1266 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1325  
Db 1717 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1776  
Qy 1326 AGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1385  
Db 1777 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1836  
Qy 1386 GGTCAAGAGCAACCAAGTATGCTGCTGTTGTCAGATTAATGGAAGCAATTA 1445  
Db 1837 GGTCAAGAGCAACCAAGTATGCTGCTGTTGTCAGATTAATGGAAGCAATTA 1896  
Qy 1446 CGATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1505  
Db 1897 TGAATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1956  
Qy 1506 CCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1565  
Db 1957 CCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2016  
Qy 1566 GAACTGCAAGATGGAAGAGAGCTTGAAGGTTGAAGATCCCTGGAAGCTGATGAT 1625  
Db 2017 GAACTGCAAGATGGAAGAGAGCTTGAAGGTTGAAGATCCCTGGAAGCTGATGAT 2076  
Qy 1626 CGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1685  
Db 2077 CGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2136  
Qy 1686 TGCCAGATCAATGAAGAGATGACCCGCTGTTCTCAAGCTGATTAATTCAGATGAGG 1745  
Db 2137 TGCCAGATCAATGAAGAGATGACCCGCTGTTCTCAAGCTGATTAATTCAGATGAGG 2196  
Qy 1746 ACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1805  
Db 2197 ACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2256  
Qy 1806 CTGCAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1865  
Db 2257 CTGCAATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2316  
Qy 1866 GCTGAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1925  
Db 2317 GCTGAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2376  
Qy 1926 TAGAGTTAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1985  
Db 2377 TAGAGTTAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2436  
Qy 1986 TTTTGTCAAGTCTGAGAGAAAGTCAAGATCCATCTCTGGAACATTAATGATGAG 2045  
Db 2437 TTTTGTCAAGTCTGAGAGAAAGTCAAGATCCATCTCTGGAACATTAATGATGAG 2496  
Qy 2046 GGTACAGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2105  
Db 2497 GGTACAGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2556  
Qy 2106 TGTTCCTCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2165  
Db 2557 TGTTCCTCAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2616  
Qy 2166 ACAGCGTTAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2225  
Db 2617 ACAGCGTTAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2676  
Qy 2226 TCTGCTCTTGAAGAGCAAGCTGCTTCTTCTGTTGAGCAAGATTAATCAAG 2285  
Db 2677 TCTGCTCTTGAAGAGCAAGCTGCTTCTTCTGTTGAGCAAGATTAATCAAG 2736









QY 126 CTTATATCATGATCATCTTGAATTAACCCAGTATGAGAAAGCCAGAAAGCTCCGTTCAAG 185  
DB 840 TGCATATTAATTAACAGCTTATGATTAACCAAGATTAAGAGATTAAGATTAATCCGTTCAAT 899  
QY 186 TCTCTCTTTTCAACACAGAGATCTAATTTGAAAGTGTCAATGCTTTGATGAAAGATTAAT 245  
DB 900 TTTACTTGATCAATGAAAGAAATTTGAGAGCCGATATATTTGATGAAAGATCTTTT 959  
QY 246 ATTTTTCCTTAAGATTAACAGCAAAAGTTTCTGAAGTTTATGAAAGCCGGAATG 305  
DB 960 ATTTTTCCTGAGCTTAAGAAAGCGAGAGTGAATGTTGAGCAAAACAAACAA 1019  
QY 306 AGAGAGTGTGAGATTAAGATCACTTTAAGAAATGAACCTTCACTATCAACCAATTTG 365  
DB 1020 AATCATGCTGAAGTTTACAGTTGATTTTCCAAAGAACTACCGCCACGTCGCAATG 1079  
QY 366 TTTGCAAGTTCTAATTAATTTATTTTCAAGAGCTTTGAAATCATGAATTTGCAAAAT 425  
DB 1080 CCTACGCTATTAACAACTTCTCTTGAAGAACTTACAGCTGCTGATTTGAACAAAT 1139  
QY 426 TGAAGAAATTAATTAACCAATGAACCAATGATAT-----TCAAGTCAAGAT 479  
DB 1140 TGTGCGCACTATTTATTAACCAAAAGAGCCATTTCACTTATACCTGATGTTACAGTTT 1199  
QY 480 GGTGATTTGCGCTGCTTCACTACTTTCATCTTCAATGAAAGAAAGCAATCATGCTCTG 539  
DB 1200 GGAATCTGGCTGGTTATGTTACTTCTGTTCTTCAATAGCAAAACAGATTTACCTCTG 1259  
QY 540 CACTGAGCTTGAATTAAGTCTTGAAGTGAAGTCTTGAATTTGATTTCAATGTTCACTT 599  
DB 1260 TGCAGATGAGCCAAAGTCTCCGATTAAGAACTGCTTATATTTCAATTAAGAAAC 1319  
QY 600 TTATCATGACAGAGAAACATTAATTTCAAGAACTGTTCCAAAGAACTAATAGTTT 659  
DB 1230 ATCTGCCAGGCCAGACAGAAACATCCAGAGAAATCACTAATTAATTTGATTC 1379  
QY 660 AGTTGTTTTCACAGTTTAACTAATTAAGTGAAGTGAATTTGAGTGGGAGCA 719  
DB 1380 AATGTTTGAACAAATTAACAACTTCAAGATTAATTTGATTTGAAGTGAAGCA 1439  
QY 720 GAATCCCAAGAGCACTTTAAGAAAGCCGAGCTGAGATGATGATTTTGAATTA 779  
DB 1440 GAATCTGAAGACATTTTAACTAATGAGTGAAGCAAAATCACTAATATAGACTTA 1499  
QY 780 CAGAGACATTAACCAAGAGATACCGACTTGAAGAGCTGCTTTGTCACAGCC 839  
DB 1500 CAGGAGACATTAAGAAATTTGTCAGATGAAGAAACAGCCACTTTTGGTCAAGCGG 1559  
QY 840 CAAGAAAGAGGGGGGCGCTGGGGGAGCACTGCGAGGGCTGCAATGCTCATTCCTGAGT 899  
DB 1560 CAGATGAGAAAGGGGCTTAAGGGTTCACAACTGAACTTATCCGCTGATTTCTCAGCT 1619  
QY 900 CTGCTATCTTACAGCTTCTACTGATTAATGCTGATATTTTAACTGATGAAGACT 959  
DB 1620 GTGCAACATGACAGCTTCTAACAGATGAATATGTAAGATTAATGATTAAGAAATTT 1679  
QY 960 AACCGTTCACTAAGACTAATCTACAGAGCAAGGAGGAGTGAAGTGAAGCACTATTA 1019  
DB 1680 GGGTAAACATACAGATTTGATGTCAGAGAAAGGAGATCATATTAAGAAATTTCA 1739  
QY 1020 TTACATTCATAAAGATTAATGTTCAAGAGAGCTTGAAGCTGGGTTGAGCTTTGA 1079  
DB 1740 TACTGTACAGATTAATTAAGTGAAGAGTACTTCACTCTGGGTTTGAATTTGA 1799  
QY 1080 TTTCACTACTGCTCTCTCTCAGAGAAATTTTGAAGCAAGAAAGATTTCAAGAGT 1139  
DB 1800 TACCAATTTTGTGCTGCTCCGAGAGAGTTTGAAGAAAGCAATCTGTCAGAGAG 1859  
QY 1140 AAAAAATTTGATTAACATCAAAATTTGAGATTTGCTCAAGAAAGCAAGAGTCAAC 1199  
DB 1860 AAGAAATGTTAAGCCAAATTTCAAA--GAGAGCTGTCAGAGAAATTAAGAAATTTCC 1916  
QY 1200 ATTAATTAAGTGAAGCACTAGATTAATGCTGTTGATTTATAGCGAAAGAAATTAAGA 1259

DB 1917 CTTACTTAATGCAATGCGACATACATAGTTGGCTCATCTATATGACAGAGACATGACAG 1976  
QY 1260 AGAGCCCAATTCATGATTAACAAATCTATTTAAGTTTACACAGCACTGGGATGCAAT 1319  
DB 1977 AGAGCCATGCTCTTAAAGGCTCATCTACAGAGTGTCAAGCCGCCCATGGCATTAATCTAT 2036  
QY 1320 GAGAAAGCAATTAATGATTAAGTGAATGACAACTGAAGCTTACCTTAAAGTCTTACA 1379  
DB 2037 GAAACCAAGAAATGATTAAGTGAATGATGATGATTAATCTTATTAAGACATTAAG 2096  
QY 1380 GCAAAAGTCAACAGAGACCCAGAT-----AGTTGT 1412  
DB 2097 GAAATATCTAGACCAACATGCAAGATGGAATGATCATGTTGCTGTTTCAAGTTAT 2156  
QY 1413 CTGCTGTTGTCAGATTAACGAAAGCAAAATAGATGCTATTAATAATTAATCTGTTAC 1472  
DB 2157 TTTGATCTGCGCAATGATGACAACTGATATGACAGCATTAATAAGATCTATGTAAC 2216  
QY 1473 AGATTGCTTACCCCAAGTCAAGTGTGTGTGTGCGCCGAACTTGAAGCAACAGCAACTGT 1532  
DB 2217 CAATGCTCAATTTCAAGCCAGTGTGTGTGTGTAAGAACTTTAGAAAGTGTCAAGCA-- 2274  
QY 1533 CATGGCAATGCTTCAAAAGATTTGCCCTACAGATGAATGCAAGATGGAAGAGCTGTG 1592  
DB 2275 -AGACATCTGTCACAAAGATTTGCCAGACAGATTAATGCAAGATGGAAGAGCTGTG 2333  
QY 1593 GAGGTGACATCTCCCTGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1652  
DB 2334 GAAGGTGAGACAGACGTAACAAAGAACTATGTTCTGTGATTAATGTTTCAAGATAT 2393  
QY 1653 GACAGTGGGGGAGAGTGAATGAGAGATTTTGTGCAAGATCAATGAAGGATGACCG 1712  
DB 2394 CCAATATCAAGAAATCAATGATGAGATTTGTGCAAGTGTGCAATGCTGAATTAACAA 2453  
QY 1713 CTGTTTCTACGCTGATATTTTCAAGATTAAGAGACAGAGCTGTGTGTGTGTGTGTGTGT 1772  
DB 2454 GTGTGACTCTCAATGATGATCACAAGAAACAGAGAAAGCTGTGTGAAGAGCTGTGAGAT 2513  
QY 1773 CTGCTGCAAGGGGCTGTGAAGGCTTGAATAGCTGCAATGATGATGATGATGATGATGAT 1832  
DB 2514 CTGCTGAAAGTGTGCTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2573  
QY 1833 CATCTGTACCGCGATGAGGATGAGAGACGCGAGCTGAAACATCTGTGTGTGTGTGTGTGT 1892  
DB 2574 TATGTGTATCGGAGTGAAGTGTGAGATGTGATGATGATGATGATGATGATGATGATGAT 2633  
QY 1893 GCGACGTTTTTGTGATGT 1952  
DB 2634 GAAAAAGATGTGACTTATTAACATCTCT--CTTAACATTTTCACTGCTGT 2690  
QY 1953 AATGTGTGAAGAAAGATGAACACAGATTTTGTCTGATGCTGAGAGAACTTTCA 2012  
DB 2691 CATTTGTGAGAGAAACGATTAACCTAGATTTTCTTAAACATGAGAGCAATTTTCA 2750  
QY 2013 GAATTCATTTCTGGAACATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2072  
DB 2751 AATTCACCTTCAGAGAACTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 2810  
QY 2073 TTTTATGTTGAGCGAGGCTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2132  
DB 2811 TTTTATGTTGAGTCAATCTGT 2870  
QY 2133 CTATGACACAGGCGCTGAAGGACACACACATGAGCGCTGTAAGCTTAAGAGTGTGTGTGT 2192  
DB 2871 CTATGACACAGATTTGTGAGCCCAATGACATGAGCGCTTTTACATATTTGTCTATGTGT 2930  
QY 2193 CATCTATTAACATGCGCAGAGTGTCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2252  
DB 2931 CATGTATTAATTTTGTGAGGATCATTCGAGATTTTCAAGGCGCTTGTCAATATCCCAAA 2990  
QY 2253 GCTGCTTTTCTTGT 2312

Db 2991 GCTGCTTACTCGTGGGCGAGTCATTCACACAGAACCAATGCTTCTGTGCACTCG 3050  
 QY 2313 CCTTACTACTCTTAA 2328  
 Db 3051 TCTCTTACTCTTGA 3066

RESULT 11  
 LOCUS BC031060 3006 bp mRNA linear PRI 06-OCT-2003  
 DEFINITION Homo sapiens piwi-like 2 (Drosophila), mRNA (cDNA clone MGC:35296  
 IMAGE:5271278), complete cds.  
 ACCESSION BC031060  
 VERSION BC031060.1 GI:21410560  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 3006)  
 Strassberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, P.S., Wagner, L., Sherman, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Nak, S.I., Wang, J., Hsieh, P.,  
 Diachenko, L., Marisina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stappleton, M., Soares, M.B., Bonaldo, M.F., Rubin, G.M., Hong, L.,  
 Scheer, T.E., Brownstein, M.J., Uedlin, T.B., Toshitsugu, T.,  
 Carrington, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McGowan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Wotey, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Heltun, E., Kettelman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bonfield, W.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smalls, D.E.,  
 Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE  
 JOURNAL MEDLINE  
 MEDLINE 12477932  
 PUBMED 22388257  
 REFERENCE 2 (bases 1 to 3006)  
 Strassberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT  
 REMARK  
 COMMENT Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshitsugu and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIND)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [medepaxil.stanford.edu](mailto:medepaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
 R. M.  
 Clome distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LIND at: <http://image.llnl.gov>  
 Series: IRAX Plate: 48 Row: b Column: 15  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 22748904.  
 Location/Qualifiers  
 1..3006  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"

Gene  
 CDS  
 /db\_xref="taxon:9606"  
 /clone="MGC:33296 IMAGE:5271278"  
 /tissue\_type="Testis"  
 /clone\_id="NH MGC\_97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 1..3006  
 /gene="HIM12"  
 /note="synonym: FLJ36156"  
 /db\_xref="LocusID:143689"  
 4..3562  
 /product="piwi-like 2"  
 /product\_start=1  
 /product\_end=2  
 /protein\_id="AAH31060.1"  
 /db\_xref="GI:21410561"  
 /translation="MSGRRVKKARIGARSPATVEGRIGQAPLPSVDLSNNEASSN  
 GFLGSRISITNDKYGISGDSAGSTWEGVKKQKDFMDISCTREKLAVRCKGSS  
 GIVKIVLNLNLDPODQOLQYHTVYIPDLASRIPLALYSSTLNKAKDGA  
 ILFLSKLEKTELSSETORGETIKMTLLKGLPSSPVCIOVNNIFKILKLS  
 MVOIGRNFYNSPMEIPQKSLMPGEALSVSYERKLSFADSVKYLRETVLEF  
 NYALCORTGLSCFTQCEKOLGLVLYERNRYSIDIDMSVXETHFORGDTEI  
 TYVDYKQOYDITVSDINQPMVLILKKRNNSEQLAHLLPELFLGLDQAVSD  
 POLMKRAVKETLSLSPSGRQRLARVDNQRNTNARFELGTWGLFGSGLSTGRVP  
 SKRLMODHICQVSAADMSKDIRCKLINAQSLNTWLICSDRETVASEFLNCLR  
 VASGMRFNVDYKRIKIVQNSPPAPRAIQOYVDPVQVLMCTLPSQKTYDSIKKYL  
 SSQCPVSCVIALRLINKQNMISITKIAMQTCLEGLMAVEIPLKSLMVGIDV  
 CDALSKDVNVGCVASVNPRTTRMPSRIQRTMDVADCLKFNTHLNNKYTNH  
 DIPARIVRAVAGDQGLKTLIEVDPOLSSVASSSTSRSLSVTVRRKCMPEF  
 TENKRVQNPPLGTIVDSATENVDYFLISQVCRGVTSPTSYLVYDDGLKRDH  
 MORLTKLCHLYNMPGIVSVPACQYAHKTLFLVAQSIHKPSLELAHLVYL"  
 817..1224  
 /note="PZ; Region: PZ domain. This domain is named PZ  
 after the proteins Pw1 Argonaut and Zwillie. This domain  
 is found in two families of proteins that are involved in  
 post-transcriptional gene silencing. These are the Pw1  
 family and the Dicer family, that includes the Carpel  
 factory protein. The function of the domains is unknown  
 but has been suggested to mediate complex formation  
 between proteins of the Pw1 and Dicer families by  
 hetero-dimerisation"  
 /db\_xref="CDD:pfam02170"  
 1639..2517  
 /note="Pw1; Region: Pw1 domain. This domain is found in  
 the protein Pw1 and its relatives. The function of this  
 domain is unknown"  
 /db\_xref="CDD:pfam02171"

misc\_feature  
 misc\_feature  
 ORIGIN  
 Query Match 29.2%; Score 680; DB 9; Length 3006;  
 Best Local Similarity 56.7%; Pred. No. 4.5e-172;  
 Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;  
 QY 1 ATGATCTTGGTGTGAACACAGGCGAAGTGAACCTTAAAGATCAAAACAGGT 60  
 Db 244 ATGATTTGGATGTCTGTACACAGAAATTTGACATGTGAATTTTAAACAGGT 303  
 QY 61 TCTTACGACATTTAGTAAGTTAAGCACTAACCATTTCCGGCTGACATCCGTCAG 120  
 Db 304 TCCAGTGAATACCTGGAACATCGTTACAAACCTTTTAACTTAATTTCCCAAGAC 363  
 QY 121 TGAGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
 Db 364 TGAGAGCTATACAGTACATGATGATGATGATGATGATGATGATGATGATGATGAT 423  
 QY 181 TGAGCTTCTTTTTCACACGAAAGATCTAATGAAAGTGTGATGTTTGAATGACG 240  
 Db 424 ATTGCTTACTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483  
 QY 241 ATATATATTTTAACTTAAAGCTTACAGCAAAAGGTTTGAAGTTTGAATGACCGG 300  
 Db 484 ATCTTTTCTGTACAAAGAGTGAAGAAAGGTGACAGAGTTGTGATGAAATGCAAA 543



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuko, Y.

TITLE  
Full-length cDNA sequences  
Patent: EP 1308459-A 1676 07-MAY-2003;  
JOURNAL  
Helix Research Institute (JP) ; Research Association for  
Biotechnology (JP)

FEATURES  
source  
1. .3138  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 29.2%; Score 680; DB 6; Length 3138;  
Best Local Similarity 56.7%; Pred. No. 4.5e-172;  
Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

QY 1 ATGATCTTTGGTGTGACACAGGCGAAGCTTGAACCATGTTAAAGAAATCAAAAAGAGT 60  
DB ATGATCTTTGGTGTGACACAGGCGAAGCTTGAACCATGTTAAAGAAATCAAAAAGAGT 60  
QY 392 ATGATCTTTGGTGTGACACAGGCGAAGCTTGAACCATGTTAAAGAAATCAAAAAGAGT 451  
DB ATGATCTTTGGTGTGACACAGGCGAAGCTTGAACCATGTTAAAGAAATCAAAAAGAGT 451  
QY 61 TCTTCAGGATTAATAGTAAAGTTAAGCACTTAACCAATTCGGCTGACATCCGCTCCGAG 120  
DB TCTTCAGGATTAATAGTAAAGTTAAGCACTTAACCAATTCGGCTGACATCCGCTCCGAG 120  
QY 452 TCCAGTGAATACCTGCTGAACCTGTTACAAACCTCTTAAGATTTTCCCAAGAC 511  
DB TCCAGTGAATACCTGCTGAACCTGTTACAAACCTCTTAAGATTTTCCCAAGAC 511  
QY 121 TGGGCTTAATACATGATACATGATCTTAACCCATGATGAAGCCGAAGACCTCGT 180  
DB TGGGCTTAATACATGATACATGATCTTAACCCATGATGAAGCCGAAGACCTCGT 180  
QY 512 TGGCACTATACAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 571  
DB TGGCACTATACAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 571  
QY 181 TCAAGCTCTCTTTTCAACAGCAAGATCTAATGGAAGTGCATGCTTTGATGAAG 240  
DB TCAAGCTCTCTTTTCAACAGCAAGATCTAATGGAAGTGCATGCTTTGATGAAG 240  
QY 572 ATTCCTTTACTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 631  
DB ATTCCTTTACTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 631  
QY 241 ATATATATTTTACTTAAAGACTACAGCAAAAGTTACTGAAGTTTATGTAAGACCGG 300  
DB ATATATATTTTACTTAAAGACTACAGCAAAAGTTACTGAAGTTTATGTAAGACCGG 300  
QY 632 ATCTTTTCTGTCAAAAAGCTAGAAAGAAAGTGCACAGATGTGCAAGTGAATCTCA 691  
DB ATCTTTTCTGTCAAAAAGCTAGAAAGAAAGTGCACAGATGTGCAAGTGAATCTCA 691  
QY 301 AATGAGAGATGCTGAGTAAGCACTTTAAACAATGAATCCACTCATACACA 360  
DB AATGAGAGATGCTGAGTAAGCACTTTAAACAATGAATCCACTCATACACA 360  
QY 692 AGAGTGAAGCTAATGAAGTACATGATGATGATGATGATGATGATGATGATGATGAT 751  
DB AGAGTGAAGCTAATGAAGTACATGATGATGATGATGATGATGATGATGATGATGAT 751  
QY 361 ACTTGTTCAGTCTTATATATATATTTTCAAGAGCTTTGAAAATCAATGAATTCGA 420  
DB ACTTGTTCAGTCTTATATATATATTTTCAAGAGCTTTGAAAATCAATGAATTCGA 420  
QY 752 GTGTGATCAAGGCTTCAATATCATCTTCAGAAAATCTCAAAAAGTGTGCAATGAC 811  
DB GTGTGATCAAGGCTTCAATATCATCTTCAGAAAATCTCAAAAAGTGTGCAATGAC 811  
QY 421 CAATGAGCAAAAT 480  
DB CAATGAGCAAAAT 480  
QY 812 CAATGAGCAAAAT 871  
DB CAATGAGCAAAAT 871  
QY 481 GTGATTTGGCTGGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 540  
DB GTGATTTGGCTGGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 540  
QY 872 TCCCTTTGGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 931  
DB TCCCTTTGGCTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 931  
QY 541 ACTGACGTTAGCAATTAAGTCTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600  
DB ACTGACGTTAGCAATTAAGTCTTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600  
QY 932 GCTGAGTGTGATTAAGAGTCTTCGGAATGAGACGTTCTGAAATTCATGCTGCTTC 991  
DB GCTGAGTGTGATTAAGAGTCTTCGGAATGAGACGTTCTGAAATTCATGCTGCTTC 991  
QY 601 TATCATCAAGCAGAGAAATTAATTTCAAGAAATTTCCAAAGATTTAATAGTTTA 660  
DB TATCATCAAGCAGAGAAATTAATTTCAAGAAATTTCCAAAGATTTAATAGTTTA 660  
QY 992 TGTCAAGAACTGGCTTGTCTCTTTCACCCAGACGTTGAGAAAGAGTATATAGGCTC 1051  
DB TGTCAAGAACTGGCTTGTCTCTTTCACCCAGACGTTGAGAAAGAGTATATAGGCTC 1051  
QY 661 GTTGTCTTCAAGATTAACAATTAAGACATAGAGTGAATATTTGACTGGGACAG 720  
DB GTTGTCTTCAAGATTAACAATTAAGACATAGAGTGAATATTTGACTGGGACAG 720  
QY 1052 ATGTCTTCAAGATTAACAATTAAGACATAGAGTGAATATTTGACTGGGACAG 1111  
DB ATGTCTTCAAGATTAACAATTAAGACATAGAGTGAATATTTGACTGGGACAG 1111  
QY 721 AATCCAGAGCACTTTTGAAGAAAGCCAGCTGAGTGAAGTGAAGTGAAGTGAAGTGA 780  
DB AATCCAGAGCACTTTTGAAGAAAGCCAGCTGAGTGAAGTGAAGTGAAGTGAAGTGA 780  
QY 1112 AAGCCACACACACTTTTCAAGAAAGCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1171  
DB AAGCCACACACACTTTTCAAGAAAGCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1171

QY 761 AGAAGCAATTAACAAGATGATCAGCACTTGAAGAGCTGTGCTGACGACAGCC 840  
DB AGAAGCAATTAACAAGATGATCAGCACTTGAAGAGCTGTGCTGACGACAGCC 840  
QY 1172 AAGCAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1231  
DB AAGCAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1231  
QY 841 AAGAGAGGCGGCGCTGAGGAGACCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCT 900  
DB AAGAGAGGCGGCGCTGAGGAGACCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCT 900  
QY 1232 AAG---AAGAGAGAAATGAACAAGAGAGCTGAGGAGCTGAGGAGCTGAGGAGCT 1288  
DB AAG---AAGAGAGAAATGAACAAGAGAGCTGAGGAGCTGAGGAGCTGAGGAGCT 1288  
QY 901 TGTCTTCTTCAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960  
DB TGTCTTCTTCAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 960  
QY 1289 TGTCTTCTTCAAGGCTGAGTGAAGCAAGCAATGATTTTCAAGTGAAGTGAAGTGAAG 1348  
DB TGTCTTCTTCAAGGCTGAGTGAAGCAAGCAATGATTTTCAAGTGAAGTGAAGTGAAG 1348  
QY 961 GCCCTGATCAAGACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
DB GCCCTGATCAAGACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020  
QY 1349 GCTGAAAAGACAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1408  
DB GCTGAAAAGACAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1408  
QY 1021 TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080  
DB TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080  
QY 1409 AACATCAAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1468  
DB AACATCAAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1468  
QY 1081 TCCAACTTACTGCTTCTTCAAGAGAAATTTGCAACAGAAAAGATTCACCAAGTGA 1140  
DB TCCAACTTACTGCTTCTTCAAGAGAAATTTGCAACAGAAAAGATTCACCAAGTGA 1140  
QY 1469 AGCCAGATA---TCTGAGCTGGCGGATTTGTGCTTCAAGAAAATTAATTAATTAAG 1525  
DB AGCCAGATA---TCTGAGCTGGCGGATTTGTGCTTCAAGAAAATTAATTAATTAAG 1525  
QY 1141 AAAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
DB AAAACATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200  
QY 1526 CACATAT---GTCAACCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1582  
DB CACATAT---GTCAACCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1582  
QY 1201 TTAATTTGTTAAAGCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
DB TTAATTTGTTAAAGCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1260  
QY 1583 ATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1642  
DB ATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1642  
QY 1261 GCAGCCAAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
DB GCAGCCAAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320  
QY 1443 GTTCCCGAGAGCTTCTGAAGCTGCTGAGAGAGTTCAGAGTTCAGAGTTCAGAGTTC 1702  
DB GTTCCCGAGAGCTTCTGAAGCTGCTGAGAGAGTTCAGAGTTCAGAGTTCAGAGTTC 1702  
QY 1321 AGAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380  
DB AGAAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380  
QY 1703 GACTACCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1762  
DB GACTACCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1762  
QY 1381 CAAGAAGTCAAGAGCAACCCAGATATGTTGTCTGTCTGTTTCAATTAATTAATTAATTA 1440  
DB CAAGAAGTCAAGAGCAACCCAGATATGTTGTCTGTCTGTTTCAATTAATTAATTAATTA 1440  
QY 1763 CAATATGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1822  
DB CAATATGTGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1822  
QY 1441 AAATACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500  
DB AAATACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500  
QY 1823 TATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1882  
DB TATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1882  
QY 1501 GTGCGCCGAACCTTAGGCAACAGCAAACTGTCAATGAGCTAATGAATTCAGCTTA 1560  
DB GTGCGCCGAACCTTAGGCAACAGCAAACTGTCAATGAGCTAATGAATTCAGCTTA 1560  
QY 1883 CTGCTGAGACCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1942  
DB CTGCTGAGACCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1942  
QY 1561 CAGATGAATCTGAAGATGAG 1620  
DB CAGATGAATCTGAAGATGAG 1620  
QY 1943 CAGATGAATCTGAAG 2002  
DB CAGATGAATCTGAAG 2002  
QY 1621 ATGATCGTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
DB ATGATCGTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680  
QY 2003 ATGATCGTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2062  
DB ATGATCGTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2062  
QY 1681 TTTGTGCGAGCATCAATTAAGAGAGATACCGCTGTTTCAAGCTGATTAATTCAGAT 1740  
DB TTTGTGCGAGCATCAATTAAGAGAGATACCGCTGTTTCAAGCTGATTAATTCAGAT 1740  
QY 2063 TGGCTGCGAGAGTGAATCCCAAGATACCAAGAGTGTTCCTCCGCTATTCCTTCAAGA 2122  
DB TGGCTGCGAGAGTGAATCCCAAGATACCAAGAGTGTTCCTCCGCTATTCCTTCAAGA 2122  
QY 1741 AGAGAGCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
DB AGAGAGCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
QY 2123 ACAAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2182  
DB ACAAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2182  
QY 1801 AATGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
DB AATGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860  
QY 2183 TACAAGTACATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2242  
DB TACAAGTACATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2242







QY 601 TATCATCAGACAGAAACATTAATTTCAGAAACAGTTTCCAAAGAACTAATAGCTTGA 660  
 DB 992 TGTCAAGAACTGCTGCTGCTTCAACCCAGAGCTGTGAAGAAACACATTAATAGGGCTC 1051  
 QY 661 GTTGTCTTCAAGATTAACAATAAGACATACAGAGTGAATGATTAATGACAGGACAG 720  
 DB 1052 ATTGTCCTTCAAGATTAACAATAAGACATTAATGATTAATGATTAATGATTAATG 1111  
 QY 721 AATCCCAAGACAGCTTTAAGAAAGCCGAGCTGTGAAGTCACTTTTGAATTAATGAT 780  
 DB 1112 AAGCCCAACACACCTTTTCAAGAGGGGAGTGGCACCAGATCACTAATGTGATTAATG 1171  
 QY 781 AAGAAAGATTAACAACCAAGATCAACCCGCTTGAAGAGCCCTGTGTGTGAGCCAGCC 840  
 DB 1172 AAGCAGCAGATTAATTAATGATGAGCTGAATCAAGCCCATCTGTGTGATGATTA 1231  
 QY 841 AAGAGAGAGCGGGGCTGAGGAGGAGCACTGAGGAGGCTGAGCTGATTCCTGAGCTC 900  
 DB 1232 AAG---AAGAGAGAAATGAACAAGTAGAGCTCAGCTGAGCCACCTGATACCTGAGCTC 1288  
 QY 901 TGTATCTTACAGCTTAACTGATTAATGCGTAAATGATTTAAAGTGAAGAACTTA 960  
 DB 1289 TGTCTTCAACAGGGCTGAGCTGAGCAGCAGATCTGATTTTCACTGATGAAGGCTGTG 1348  
 QY 961 GCGGTCATACAAAGCTTAAGCTGAGCAGAAAGCAGGCTGAAGTGGAGCACTGATGAT 1020  
 DB 1349 GCTGAAGAAAGACGCTTCACTGCTTCAAGGCGGAGAGGCTGAGCCAGGCTTGTGAGC 1408  
 QY 1021 TACATTCATTAACAGATTAATGTTCAAGAGGAGCTTCAAGCTGAGGCTTGAAGCTT 1080  
 DB 1409 AACATCCAGAGAAATCAAGATGCTGCTTGAATGAGAGACCTGGGAGCTGATTTTGA 1468  
 QY 1081 TCCCACTTACTGCTTCTCAGAGAAATTTTGCAGAAAGAAATGATCAACAAGTGA 1140  
 DB 1469 AAGCAATA---TCTGAGCTGCGGATGTGCTTCAAGAAATTAATTAATGCAAGC 1525  
 QY 1141 AAAACATTTGATTAACATCAACATTTTGCAGATTTGGTCCAAAGAAACAAGAGTGCACA 1200  
 DB 1526 CACATTT---GTCAAGCTGTGTCTGTGTGATGCTGTGATGAGATTAATGAGACTTGA 1582  
 QY 1201 TTAATTAAGTTAAGCCTAGATTAATCTGCTGTGTGATTAATGCGGAAGAAATTAATGA 1260  
 DB 1583 ATTTTAAATGACAGCTTGAATACCTGTTGATTTAATGAGCAGCAAGACTAATAT 1642  
 QY 1261 GCAAGCAATTCATTAATCAAAATCTAATTAAGTACACAGCCATGGGAGTGAATG 1320  
 DB 1643 GTTCCGAGAGCTTCTGAGCTGCTTGAAGAGATTTGCAAGTTCATGAGATTTAATGTG 1702  
 QY 1321 AGAAAGCAATAATGATTAAGTGAAGTGAAGAGAGTGAAGAGTGAAGTGAAGTGAAG 1380  
 DB 1703 GACTACCCCAATTCATTAAGTGAAGAGAGAGAGTGAAGAGTGAAGTGAAGTGAAG 1762  
 QY 1381 CAAAGAGTCAAGAGACACCAAGATGATGCTGTGTGATTAAGTGAAGTGAAGTGAAG 1440  
 DB 1763 CAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1822  
 QY 1441 AATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
 DB 1823 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1882  
 QY 1501 GTGCGCCGAGCTTGAAGCAACGCAACGCTGATGAGCAATGCTTAAGTGAAGTGAAG 1560  
 DB 1883 CTGCTGCGAGCTTGAATTAACAGGAGATGATGATGATGATGATGATGATGATGATG 1942  
 QY 1561 CAGATGAATGCAAGATGAGAGAGAGCTGCTGAGAGGAGTGAAGTGAAGTGAAGTGAAG 1620  
 DB 1943 CAGATGAATGCAAGATGAGAGAGAGCTGCTGAGAGGAGTGAAGTGAAGTGAAGTGAAG 2002  
 QY 1621 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
 DB 2003 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2062  
 QY 1681 TTGTTGCGAGCATGATGAAGGATGACCCGCTGCTGCTGATGATGATGATGATGATG 1740

DB 2063 TGGCTGAGCAGTGTAAACCCAGAAATCAAGAGTGTGTTTCCGCTGTATCTTCAGAGA 2122  
 QY 1741 AGAGAGCAGAGAGCTGTGATGATGAGGCTCAAAAGTGTGCTGAGCAGAGGCTGAGGGCTTGG 1800  
 DB 2123 ACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2182  
 QY 1801 AATAGCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
 DB 2183 TACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2242  
 QY 1861 GGCCAGCTGAAAACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920  
 DB 2243 GGTGAGCTGAGAAACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 2302  
 QY 1921 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980  
 DB 2303 TCCGCTCAATTAACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2362  
 QY 1981 AGATTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040  
 DB 2363 CGATTCCTTAACCAATGAAACCGCATGATGATGATGATGATGATGATGATGATGATGATG 2422  
 QY 2041 GTAGAGTTCACAGACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100  
 DB 2423 TCGAGAGCAACAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2482  
 QY 2101 GGTAGTGTTCCTCCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 2160  
 DB 2483 GGAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2542  
 QY 2161 CACATACAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220  
 DB 2543 CATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2602  
 QY 2221 CGTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280  
 DB 2603 AGTGTCCAGACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2662  
 QY 2281 CACAGAGGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
 DB 2663 CATTAAGACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2720

RESULT 14  
 AB079366  
 LOCUS 3158 bp mRNA linear PRI 02-AUG-2003  
 DEFINITION Homo sapiens PIV1L1 mRNA, complete cds.  
 ACCESSION AB079366  
 VERSION AB079366.1 GI:33411129  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 Sasaki, T., Shiohama, A., Minoshima, S., and Shimizu, N.  
 Identification of eight members of the Argonaute family in the  
 human genome  
 Genomics 82, 323-330 (2003)  
 2 (bases 1 to 3158)  
 JOURNAL Shimizu, N., Minoshima, S., and Sasaki, T.  
 Direct Substitution  
 Submitted (04-FEB-2002) Nobuyoshi Shimizu, Keio University School  
 of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo  
 160-8582, Japan (E-mail:shimizu@med.keio.ac.jp,  
 Tel:81-3-3351-2370, Fax:81-3-3351-2370)  
 Location/Qualifiers  
 1..3158  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"

FEATURES

source

```

gene
1..3158
/seq="PIW11"
CDS
172..2730
/seq="PIW11"
/codon_start=1
/protein_id="BAC81341.1"
/db_xref="GI:33411130"
/translation="MSGARVAVGARSPEAREVERIQASPLPREVDLSNEMASSN
GFLGSTRINDKYGISGDAGSTFERGVKNODFEDLICTREKILAHKRNKASS
GIPKLVTLNFDLPDMDQYQYHVIYLPDLSRRLIMLVSHSLSKARAFDA
ILFPOKLEEVETSESTGERTIKMTITLRELPESSPVICQFENIIPKLIKLS
MYOGRNVESEMEIPQHLISLMPGPAISVSEKELPSAASVXYLVNGLVEF
MTALCORGLSCPTOCCEKQILGLIVLRRNNRYSIDIDIMSVKPTHEOKRDEI
TVVDYKQYDITVSDLNQFVLVSLKKRDNSEBOLAHILPEICETGLDQATSD
FOLKAVAEKRLSPSGQRLARLVNDIQNTNARELEHGLHPSQSLITGRIVP
SEKILMODHIQPVSAADWSIDIRCKILNOSLNTWLICSDDEYVASFLNCLR
VAGSNGENVDPKIIKQENPAFVRAIQYVDPDVLQVLCIPESNKTYDSIKYLV
SSDCVPSQCVLARLANKQGMMSIATKIAMQTCIKGELMAYEIPKSLMVGIDV
CKDLSKDVVVGVASVNPRIITRMFSCLIORMTVDADLKFMTGALKNTKYKH
DLPARIVYRAGVDGQIKTLIEYVQLSSVAESSNTSLSVIVRKKCPREF
TEMRIVONPPLGVDSAEARNEMVDPYLSOYACGTSPTVYVINDGLKPDH
MORLEFKLHLIYMPGLVSVPAEQIAHKLTFVAOSIHKEPSLEIANHLFI"

```

## ORIGIN

```

Query Match      29.2%; Score 680; DB 9; Length 3158;
Best Local Similarity 56.7%; Pred. No. 4.5e-172;
Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

```

```

QY 1 ATGATCTTGGTGTAACAAGCAAGCAACTTGAACCTGTTAAAGATCAAAAACAGT 60
DB 412 ATGATTTGAGATCTGTACAGAGAAAATGGACATGTGAGAAATGTAAAGCAGT 471
QY 61 TCTTCAGGCAATATAGTAAGTTAGCACTAACATTCGGCTGACATCCCGTCCAG 120
DB 472 TCCAGTGAATACCTGTGAAAAGTGTACAAACCTTTAACTTGAATTTCCCAAGAC 531
QY 121 TGGGCTTATATCACTATCACTTCACTTAAACCACTGATGAAACCAAGAACTCCGT 180
DB 532 TGGCAGCTATACCACTACATGTGACATATATCCAGATTTAGCATGAGAGGTGAGA 591
QY 181 TCAGCTCTCTTTTCAACAGAAATCTAAATTGAAAGTGCAGCTTTGATGAGAAAG 240
DB 592 ATTGCTTATCTTATAGTATAGTAACTTCCAAACAGCAAAAGCATTTGACGATGC 651
QY 241 ATATTATTTTAACTTAAAGCACTAGCAAAAGGTTACTGAAAGTTTATGAAAGCCGG 300
DB 652 ATCTTTTCTGTACAAAGCTAGAAAGAAAGGTCAACAGATGTCTCAAGTGAATCTAA 711
QY 301 AATGAGAGAGATGTAGAGTAAAGATCACTTAAACAATGAACTTCCACCTACATCAACA 360
DB 712 AAGAGTGAACATAAAGATGACTATACCTGTAAGAGGAGAGCTCCATCAAGTTCTCC 771
QY 361 ACTTGTTGCACTTATATATATTTTCAAGAGGCTTTGAAAATCATGAAATTTGCAA 420
DB 772 GTGTGATCAAGGTCTTCAATATCACTTCAAGAAAGATCTCAAAAAGTTGCAATGAC 831
QY 421 CAATTTGAGCAAAATTTATTAACCAAAATGACCCCAATTTGATTTCCAAAGTACAGGTG 480
DB 832 CAATTTGAGCAAACTTCTATATCTTCAAGAGCCAAATGAAATTTCCCAAGCAAAATTA 891
QY 481 GTGATTTGGCTGGCTTCACTACTTCCATCTCTTCAATGATGAAAAAGCATCATGCTGAC 540
DB 892 TCCCTTTGGCTGGCTTCCATTTCTGTGCAATTTTGAAGAGAGTCTGTTAGT 951
QY 541 ACTGAGCTTACCCATAAAGTCTTGAAGTGAAGCTGTTTGAATTTCAATTTCAACTTT 600
DB 952 GGTGATGTGATTAACAAGTCTTCCGAAATGAGACGGTCTGGAATTCATGACTCTCTC 1011
QY 601 TATCATCAGACAGAAAGCAATAATTTCAAGAAAGAGTTTCCAAAGAACTTAATAGATT 660
DB 1012 TGTCAAAAGACTGCTTCTCTTTCACCCAGACGCTGTGAGAGAGCTAATATAGGCTC 1071

```

```

QY 661 GTTGTCTTACAGATTAACAATAAGACATACAGAGTGAATGATTTGACTGGACAG 720
DB 1072 ATTGTCTTACAGATTAACAATAAGACATACAGAGTGAATGATTTGACTGGACAG 1131
QY 721 AATCCCAAGAGCACTTTAAGAAAGCCGAGCTTGAAGTCACTTTCTTGAATATCTAC 780
DB 1132 AAGCCCAACACACTTTCAAGAGGAGGATGACACGAGATCACTAATGTGATTAATAC 1191
QY 781 AAGAGCAATACACCAAGATACACCACTTGAAGAGCCGCTGTGTCAGCCAGCC 840
DB 1192 AAGCAGCAATATGATTAATCTGATGAGACTGAATACCACTGTTTATGTTGTTA 1251
QY 841 AAGAGAGCGGGGCGCTGGGGGAGACATGCCAGGGCTGCGCATCTCATTTCTGAGCTC 900
DB 1252 AAG--AAGAGGAAATGACAAACAGTGAAGCTGCGCCACCTGATCTGAGCTC 1308
QY 901 TGTATCTTACAGCTCTAATCTGATTAATAGCGTAATGATTTTAAAGTATGAAGACTTA 960
DB 1309 TGTCTTCTAACAGGCTGATCTGACAGGCAATCTGATTTTCAAGTATGAAGCTG 1368
QY 961 GCCGTTATACAAAGCTTAACTCCAGAGCAAGCGGTAAGTGGAGCACTGAT 1020
DB 1369 GCTGAAAGACAGCTGATCTGATCTGAGCGGAGACGCGCTGGCAGGCTTGGAC 1428
QY 1021 TACATTCATTAATAACGATATATGTTCAAAGGAGCTTCAAGACTGGGGTTTGAAGTTGAT 1080
DB 1429 AACATCCAGAGAAATACCAATCTGCTTGTGAAGTGAAGACTGGGGAGCTGATTTGGA 1488
QY 1081 TCCACTTACTGCTCTTCAAGAAATTTTGAACAGAAAGATTTCAACAGAGTGA 1140
DB 1489 AGCCAGATA---TCTGTGACTGGCGGATTTGGCTTGAAGAAATATTAATGCAAGAC 1545
QY 1141 AAAACATTTGATTAACAATCAATTCATTTGAGATTTGCTCAAGAAACAAGTGAACA 1200
DB 1546 CACATAT---GTCAACCTGTGTCTGCTGCTGATCTGATCAAGAGATTTGAACTTGAAG 1602
QY 1201 TTAATTAGTTTAAAGCAGTAACTGATGATCTGATCTTATACGGAAGAAATTAAGA 1260
DB 1603 ATTTTAAAGCAGAGCTTGAATTAATCTGATTTTATGATGAGCAGAACTGATAT 1662
QY 1261 GCAGCAATTCATTTGATTAACAAATCTATTTAAAGTTAACACAGCCATGGGATGCAAAAG 1320
DB 1663 GTTCCGAGAGCTTCTGAAGCTGCTTGAAGAGGTTGAGGTTCCATGGAATTAATG 1722
QY 1321 AGAAAGCAATATGATTAAGTGAAGTGAAGAGCTGAAGAGCTTAAAGAGCTTACAG 1380
DB 1723 GACTACCCCAAAATCAATAAAGTGAAGAAATCCAGCTGCAATTTGTAAGCTATACAG 1782
QY 1381 CAAAAGTCAACAGACACACCAAGATGTTGTCTGTCTGTTGTCAAGTAATCGAAAGAC 1440
DB 1783 CAATATGATCTGATGATGTTCAAGTGAATGATGATTTCTGCTTCTATCAAGAGAC 1842
QY 1441 AAATGCAATCTTAAATAATCTGTGTAAGATTTGACCTTCAAGCTCAAGTCTGAT 1500
DB 1843 TATTATGATTCATTAATAATTTGAGCTGCAAGCTGCAAGCTTCAAGCTCAAGTCTG 1502
QY 1501 GTGCGCCGAACCTTGAAGCAACAGCAACCTGATGAGCAATTTGCTCAAAAGATTCCTTA 1560
DB 1903 CTGTCTGAGACTTGAATTAACAGGCAATGATGATGATGATGATGATGATGATGATG 1962
QY 1561 CAGATGAATGCAAGATGGAAGAGAGCTTGAAGAGTGAAGTCCCTGGAAGCTGAT 1620
DB 1963 CAGATGAATGCAAGCTGGAAGAGAGCTGATGAGAGTGAAGTCCCTGGAAGCTGAT 2022
QY 1621 ATGATCTTGAATGATTTGATTAACATGATGATGATGATGATGATGATGATGATGAT 1680
DB 2023 ATGATCTTGAATGATTTGATTAACATGATGATGATGATGATGATGATGATGATGAT 2082
QY 1681 TTTGTTCCAGATCAATGAAGAGTGAAGCTGCTTCAAGCTGATTAATTTCAAGAT 1740
DB 2083 TGCCTGAGCAGGTTTAAACCCAGATCAACAGGATGATTTCCGCTGATCTTCAAGAG 2142
QY 1741 AGAGCAGAGAGCTGATGATGAGGCTCAAGGCTGCTGCAAGCGGCTTGAAGGCTTGG 1800

```

Df		2143	ACAAATGACTGATGTTCAGATTGGCTTGAAAAGTTTTCATGACTGGAGAAGCACTCAACAATGG	2202
OY		1801	AATAGCTGCATAGATCATGCCCGCAGCCGATCATCGTGTACCAGATGGCGTAGAAGAC	1860
Df		2203	TACAAGTACATCATGATTTGCCAGCACGGATAATGTGTACCGTGGTGGTGGGAT	2262
OY		1861	GGCCAGCTGAAAAACATCGGTGAACCTAGAGAAGGCACAGTTTGGATTGTCTAAATCC	1920
Df		2263	GGTACGCTGAAAACTATTATGATTAATGAATCCACAGCTGCTGAGAGAGTGGCAGAA	2322
OY		1921	ATTGTGAGAGGTTCACACCCTAGACTAACGGTAAATGTGTGGAMAAAAAGAGTGAACAC	1980
Df		2323	TCCAGCTCAAAATACCAAGCTCAAAGCTGTGGGTGATGTGGTCAGAGAAAGTGCATGCCA	2382
OY		1981	AGATTTTTTGGCTCAGTCTGAGAGAAAGACTTCAGAAATCCACTTCCGGAACAGTTATGAT	2040
Df		2383	CGATTCTTTACCGAATAGAACCGCACTGTACAGAACCCCCCACTTGGCAGCTGTGTGAT	2442
OY		2041	GTAGAGGTTCACAGCACAGAAATGTATGACTTTTATCGTAGCCAGGCTGTGAAGT	2100
Df		2443	TCAGAGAGCAACAGTAAGAAATGTATGACTTTTATCGATACGCCAGGTGGCCGCGG	2502
OY		2101	GGTAGTGTTTTCTCCACACATTACATGTCTATAGACACAGGGCGCTGAAGCACAAC	2160
Df		2503	GGAACTGTAGTGGCTTACCTACTATATGTATGTCTATATGATGACMAAGGCTTGAAGCCGAC	2562
OY		2161	CACATACAGCGCTTACCTACCAAGCTGTGCCACATCTATTACATGCGCCAGTGTCAAT	2220
Df		2563	CHATGACGAGAACTTACATCAAAATGTGTGCCACCTGTACTACAACTGGCGGGCATATGC	2622
OY		2221	CGTGTCTCTGCTCCTTGCACAGTACGCCCAAGCTGGCTTTTCTGTGGCAGAGTAT	2280
Df		2623	AGTGTCCCAAGCACCATGTCAATGATCTCAACAAGCTGAACTTCTGTGTGGCACAAGACAT	2682
OY		2281	CACAGAGAGCCAAATCTGTCACTGTCCAACCGCTTACTACCTCAAA	2328
Df		2683	CATTAAGAACCCAGTCTGGAATTAGCCACACATCTCTTCACTGTGA	2730
RESULT_15				
AX834401			2456 bp DNA linear PAT 15-DEC-2003	
LOCUS				
DEFINITION			Sequence 1525 from Patent EP1347046.	
ACCESSION			AX834401	
VERSION			AX834401.1 GI:39920536	
KEYWORDS				
SOURCE				
ORGANISM			Homo sapiens (human)	
REFERENCE			Homo sapiens	
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Huteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hiro,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuko,Y.	
TITLE			Full-length cDNA sequences	
JOURNAL			Patent: EP 1347046-A 1525 24-SRP-2003;	
FEATURES			Research Association for Biotechnology (JP) Location/Qualifiers 1..2456 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"	
ORIGIN				
Query Match			26.5%; Score 616.8; DB 6; Length 2456;	
Best Local Similarity			57.5%; Pired. No. 5,2e-155;	
Matches 1170; Conservative			0; Mismatches 857; Indels 9; Gaps 3	
OY		293	AGACCCGGATGAGAGGATGTGAGATTAACATCATCTTAAACAATGAATCCACCTA	352
Df		2	AATCTCAAAAGGTGACATTAATAATGACTTCAACCTCGAAGAGAGAGCTGCATCA	61

QY	355	CATCACCAACTGTTGGAGCTTCATATATATATATTTTCAGGAGGCTTTGAAATCATGA	412
Db	62	GTTCCTCCGTGTGATCCAGGTCTTCATATCATCTTCAGAAAGATCTCCAAAAGTTGG	121
QY	413	ATTGCAACAAATTGACGCAATTTATTAATACCAGAAATGACCCAAATGATATTCGAATC	472
Db	122	CCATGTACCAAAATGAGCGGACCTTATATATCTTCAGAGCCAAATGGAATTTCCGACG	181
QY	473	ACAGGTGATGATTTGGCCCTGAGCTCAACATCTCCATCTCTTCAGATGTAAGAACAGATCA	532
Db	182	ACAAATTAATCCCTTTGGCTCGGGTTGGCAATTTCTGTGTATATTTTGAAGAGAGCTCC	241
QY	533	TGCTGTGACTGAAGTTAGCCATTAAGTCTTTGGAATGAGACTGTTTGGATTTCAATG	592
Db	242	TGTTATGTGTGATGTGATTAACAAAGTCTCCGGAATGAGAGGTTCTGGAAATTCATGA	301
QY	593	TCAACTTTATCATTCAGACAGAGAAACATAAATTTCAAGAACAGTTTCCAAAGACTAA	652
Db	302	CTGTCTCTGTCAAGAACTGGCTGTGTCTGTTTCAACGAGGTGTGAGAACAGACTAA	361
QY	653	TAGGTTTAGTTGTTCTTAACAAGTATACATTAAGACATTCAGAGTGAATATTTAGCT	712
Db	362	TAGGGCTCATGTGCTCTTAACAAGTATACATTAACGAACCTATCTCATTTGAAGACATTA	421
QY	713	GGAGACGAATATCCCAAGACACTTTAAAGAACCGGCTCTGAAGTCAAGCTTTCTAG	772
Db	422	GGTACGTGAAGCCCAACACATCTTTTAAAGGGAGATGGCAACGAGATCACTATGTGG	481
QY	773	AATATCTACAGAGCAATACACACAGAGATCACCGACTTGAAGCAGCTGTCTGTGCA	832
Db	482	ATTATCTACAGAGCAATATGATATTTATCTGTATCGACTGAATACAGCCCAATGCTTTA	541
QY	833	GCAGCCCAAGAGAGCGGGCCCTCGGGGGACATCTGCAGGGCTCTGCCATGCTCATTC	892
Db	542	GTCTGTTAAAC--AAGAAAGAAATGACAACTGATGCTCATCTGGCCCACTGATAC	598
QY	893	CTGAGCTCTGCTATCTTACAGGTCTAATCTGTAAATATGCTAATGATTTTAACGTATGA	952
Db	599	CTGAGCTCTGCTTTCTTAACAGGGCTGTGACTGACGAGCACTGTGATTTTCAAGCTGATGA	658
QY	953	AAGACTTACCGCTTATACAGACTAATCTCAAGACAAAGCAGCGTGAATGAGACAC	1012
Db	659	AGGCTGTGCTGTAAGAAAGACGTCTCAGTCTTCAAGGCCGAGCAGCGCTGSCCAGGC	718
QY	1013	TCATTGATTACATTCATTAACATATATGTTCAAAGGAGCTTCGAGACTGGGGGTTGA	1072
Db	719	TTGTGACAAATCATCAGAGAAATCACAATGCTCCCTTGAATCTAGAGACCTGGGGATCTG	778
QY	1073	GCTTGTATTCAACTTACTGCTCTTCAGAGAAATTTTGCAACAGAAATGATTCACC	1132
Db	779	ATTTTGGAGCAGATA---TCTGTGATGGCCGAGATGTGCTCTTCAGAAATTAATTTAA	835
QY	1133	AAGGTGAAAAACATTGATTTCAATCAACAATTTGACAGATTTGGTCCAAAGAACAGAG	1192
Db	836	TGCAAGACCAATAT--GTCAACCTGTGTGCTGTGACTGTGCTCCAGGATTTTGGA	892
QY	1193	GTGACCAATTAATAGTGTAAAGCCACTAGATACTGGCTGTGATCTATACGGGAAGAA	1252
Db	893	CTTGCAAAATTTTAATGACAGCTTTGATATCTGGTTGATTTTATGTGACACAGAA	952
QY	1253	ATTATGAAGACCCCAATTCATTGATACAAAATCTATTTAAATGTACACAGCATGGGCA	1312
Db	953	CTGATATATGTTGCCAGAGCTTTCTGAATCTGCTGAGAGAGTTACAGCTTCATGGAT	1012
QY	1313	TGCAATATAGAAAAGCAATATATGATTGAAGTGAATGACAGAACTGAACCTACTTAAGAG	1372
Db	1013	TTTAATGTGACTACCCCAAAATCATTAAGATACAGAAATATCCAGCTCATTTGTTAAG	1072
QY	1373	TCTTAACGAAAAGGTCAACAGACACCCGATATGTGTCTGTCTGTGCTCAAGTATAC	1433
Db	1073	CTATACACCAATATGTGATCTGTGATGTTCAGTGTGATATGTGCAATCTGCTCTTATAC	1133

Wed May 12 09:50:20 2004

us-10-043-774b-1.rge

Page 24

QY	1433	CGAAGGCAAAATACGATGCTATTAATAAATACCTGCTGTACAGATATGGCCATACCCCAAGTC	1492
Db	1133	AGAAAGACCTTATATGATTCATTTAAATAATTTTAGCTCAGACTGCCAGTCCCAAGCC	1192
QY	1493	AGTGTGTGTGGGCCCGCAACCTTTAGCCAAACAGCAACCTGTCAATGGCATTGTCAAAAGA	1552
Db	1193	AATGTGTGCTTGTCTGGACCTTTGATTAATAACAGGGCAGTAGATGATAGATCGCAACCAAGA	1252
QY	1553	TTGCCCTTAACAATGCACTGCAAGATGGGAGAGAGCTCTGGAAGGTGGGACATGCCCTCGA	1612
Db	1253	TGCGTTATGCAAGTACCTTGCAAGGTCCGAGGCCGAGCTGTGGGCTGTGGAAATACCTTTAA	1312
QY	1613	AGCTCGATATATGCTGTGGCATCATCATTTGTTACCATGACATGACAGCTGGGCGGAGGTCAA	1672
Db	1313	AGTCCCTGATGTGTGTGTGCTGCTATATGTCTGTAAAGATGCACTCAGCAAGGACGTATGG	1372
QY	1673	TCGACAGATTTGTTGCCAGCATCAATGAAGGATGACCCCGTGGTCTTCACGCTGCATAT	1732
Db	1373	TTGTGTGATGTGGTGGCCAGTGTTAACCCCAAAATCACCGAGTGATTTTCCCGCTGATATCC	1432
QY	1733	TTTCAGATAGAGACAGAGAGCTGGATGATGGGCTCAAAAGTGTGGCTCCAAAGCGGCTCTGA	1792
Db	1433	TTTCAGAGAACATGACTGATGTGGCAGATTGCTTGAAAGTTTTCATGACTGAGACATCA	1492
QY	1793	GGGCTTGGAAATAGCTGCATAGATGATCATGCCCAAGCCGAGTATCGTGTATCCGCGATGGCG	1852
Db	1493	ACAAATGGTACAAAGTACAATCATGATATTTGCCAGACCGGATTAATGTGATCCGTGCTGATG	1552
QY	1853	TAGAGAGACGGCCAGCTGAAACACCTGGGGAATCTACGAAGTSCCAAGTTTGGATGTCTC	1912
Db	1553	TAGGGGATGGTCAAGTGTAAACACTTATTTGAATTAAGTCCCAAGCTGCTGAGCAGTGTG	1612
QY	1913	TAAATCATTTGTGTAGAGTTCACACCTTAGCTTAACGGTAAATGTGTGTGAAGAAAGAG	1972
Db	1613	TGGCAGAAATCCAGCTCAAAATACCAAGCTCAAGACTGTGCGTATGTGTGTGTCAGAGAAAGAT	1672
QY	1973	TGAACACCAAGATTTTGTGTCAGTCTGAGAGAAAGCTCAAAATCCATTCCTCTGGAACAG	2032
Db	1673	GCATGCCAGATTCCTTACCGAAATGAACCGCACTGTACAGAAACCCCACTTGGCATGTG	1732
QY	2033	TTATTTGATGTAGAGTTTACAGAGCCAGATGGTATGACTTTTATTCGTGAGCCAGGCTGTG	2092
Db	1733	TTGTGTGATTCAGAACCAACACACTAAGATGTATGATGACTTTTATGTATGAGCCAGGTGG	1792
QY	2093	TGAGAAGTGTAGTGTCTTCTCCCAACATTTACATGATCATCTATGACAAACAGCGGCTCTGA	2152
Db	1793	CTGTGCGGGGGAACGTGTAGTCTTACCTTATATATGATCATCTATATATGACAAACGGCTTGA	1852
QY	2153	AAGCCAGACCAATACAGGGCTTGAACCTTCAAGCTGTGCAACATATTTACAACTGGCCAG	2212
Db	1853	AGCCGCAACATATGAGAGACTTACATTTCAATTTGTGCAACCTGTATCTACAACTGGCCGG	1912
QY	2213	GTGTCAATTTGTGTCTCTGTCTTCCAGTACGCCCAACAGTGGCTTTTCTGTGTGGCC	2272
Db	1913	GCATATGTCAGTCTCCACAGACCATGTCAAGTATGTCTCACAAAGTGAACCTTTCGTGGGCAAC	1972
QY	2273	AAGATTTTACAGAGAGCAATATCTGTCACTGTCAAAACCGGCTTTACTACTCTTAA	2332
Db	1973	AAGGCACTTATTAAGAACCAAGTCTGGAATTAGCCAAACCATCTCTTACTACTGTGA	2032

Search completed: May 8, 2004, 14:09:03  
Job time : 9083 secs

[illegible]

Wed May 12 09:50:21 2004

us-10-043-774b-1.rst

Page 2

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source 1.2586  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/locus\_tag="HGM3939"

gene  
locus\_tag="HGM3939"

Query Match 99.5%; Score 2316.2; DB 29; Length 2586;  
Beet, Local Similarity 99.9%; Pred. No. 0;  
Matches 2318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGTGTGAACACAGGAGCACTTAAGCCATGTAAAGATCAAAAAAGATTCTTCAG 67  
DB 266 TTGGTGTGAATACAGGAGCACTTAAGCCATGTAAAGATCAAAAAAGATTCTTCAG 325  
QY 68 GCATTATATGTAAGTTAAGCACTTAACCTTCCGGCTGACATCCCGTCCCAATGGGCTT 127  
DB 326 GCATTATATGTAAGTTAAGCACTTAACCTTCCGGCTGACATCCCGTCCCAATGGGCTT 385  
QY 128 TATATCAGTATCAGTATGATTAACCACTGATGAAAGCCAGAAAGCTCCGTTCAGCTC 187  
DB 386 TATATCAGTATCAGTATGATTAACCACTGATGAAAGCCAGAAAGCTCCGTTCAGCTC 445  
QY 188 TTCTTTTTCACACGGAAGATCTAATGTGAAGAGTCTATGCTTTTGAATGAACATATAT 247  
DB 446 TTCTTTTTCACACGGAAGATCTAATGTGAAGAGTCTATGCTTTTGAATGAACATATAT 505  
QY 248 TTTTAACCTTAAGAACTACAGCAAAAGTTACTGAAGTTTATGTAAGCCGGAATGAG 307  
DB 506 TTTTAACCTTAAGAACTACAGCAAAAGTTACTGAAGTTTATGTAAGCCGGAATGAG 565  
QY 308 AGATGTGAGATTAACGATCACTTAACCAATGAACCTTCACTACATCAACCACTTGT 367  
DB 566 AGATGTGAGATTAACGATCACTTAACCAATGAACCTTCACTACATCAACCACTTGT 625  
QY 368 TGCAGTTCTAATATATTTTTCAGAGAGCTTTTGAAGAAATGAAATTTGCAAAATG 427  
DB 626 TGCAGTTCTAATATATTTTTCAGAGAGCTTTTGAAGAAATGAAATTTGCAAAATG 685  
QY 428 GACGAATTTATTAATACCAATATGATTCGAAAGTCAACAGTTGTGATTT 487  
DB 686 GACGAATTTATTAATACCAATATGATTCGAAAGTCAACAGTTGTGATTT 745  
QY 488 GAGCTGGCTTCATCTTCATCTTCATGATTAAGAAAGATGATGCTGCACTAGC 547  
DB 746 GAGCTGGCTTCATCTTCATCTTCATGATTAAGAAAGATGATGCTGCACTAGC 805  
QY 548 TTAGCATTAAGTCTTCGAAAGTGAAGCTGTTTGAATTCATGATTCATCTTATCATC 607  
DB 806 TTAGCATTAAGTCTTCGAAAGTGAAGCTGTTTGAATTCATGATTCATCTTATCATC 865  
QY 608 AGACAGAAACATTAATTTTCAAGAAAGTTTCCAAAGAACTAATAGTTTATTTGTC 667  
DB 866 AGACAGAAACATTAATTTTCAAGAAAGTTTCCAAAGAACTAATAGTTTATTTGTC 925  
QY 668 TTACCAAGTATTAATTAAGACATACAGAGTGAATGATGATGATGATGATGATGATG 727  
DB 926 TTACCAAGTATTAATTAAGACATACAGAGTGAATGATGATGATGATGATGATGATG 985  
QY 728 AGAGCACTTTAAGAAAGCCGAGCTGTAAGTGAAGTCTTCTTGAATTAATCAAGAAC 787  
DB 986 AGAGCACTTTAAGAAAGCCGAGCTGTAAGTGAAGTCTTCTTGAATTAATCAAGAAC 1045  
QY 788 AATACCAACCAAGATCAACCACTTGAAGCAAGCTGTCTTGTGTCAGCAGCCAGAGAA 847  
DB 1046 AATACCAACCAAGATCAACCACTTGAAGCAAGCTGTCTTGTGTCAGCAGCCAGAGAA 1105

QY 848 GGGGGGGGCTGGGGGGGACATGCGCAGGGGCTGGCAGTCTATTCCTGAGCTCTGATTC 907  
DB 1106 GGGGGGGGCTGGGGGGGACATGCGCAGGGGCTGGCAGTCTATTCCTGAGCTCTGATTC 1165  
QY 908 TTACAGGCTCTAATCTGATTAATTAATGCGTAAATGATTTTAAAGTGAAGAACTTACGCTTC 967  
DB 1166 TTACAGGCTCTAATCTGATTAATTAATGCGTAAATGATTTTAAAGTGAAGAACTTACGCTTC 1225  
QY 968 ATACAGACTAATCTCAAGCAAAAGCGCGTGAAGTGGGAGCATGATGATTCATTC 1027  
DB 1226 ATACAGACTAATCTCAAGCAAAAGCGCGTGAAGTGGGAGCATGATGATTCATTC 1285  
QY 1028 ATAAAAAGATTAATGTTCAAAAGGAGCTTCGAGATCGGGGTTTGAAGCTTGAATTCAACT 1087  
DB 1286 ATAAAAAGATTAATGTTCAAAAGGAGCTTCGAGATCGGGGTTTGAAGCTTGAATTCAACT 1345  
QY 1088 TACTGCTCTTCAGAGAAAGATTTTTCGAAACAGAAAGATTCACCAAGTGGAAAAAGAT 1147  
DB 1346 TACTGCTCTTCAGAGAAAGATTTTTCGAAACAGAAAGATTCACCAAGTGGAAAAAGAT 1405  
QY 1148 TTGATTAATCAATCAATTTGAGATGTCGAAAGAAACAGAGTGCACCATTAATTA 1207  
DB 1406 TTGATTAATCAATCAATTTGAGATGTCGAAAGAAACAGAGTGCACCATTAATTA 1465  
QY 1208 GTGTAAAGCACTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1267  
DB 1466 GTGTAAAGCACTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1525  
QY 1268 ATTCAATGTAACAAATCTAATTAAGTTTACACAGCATGGGCAATGCAATGAGAAAG 1327  
DB 1526 ATTCAATGTAACAAATCTAATTAAGTTTACACAGCATGGGCAATGCAATGAGAAAG 1585  
QY 1328 CAATTAATGTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1387  
DB 1586 CAATTAATGTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1645  
QY 1388 TCAAGCAAGACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1447  
DB 1646 TCAAGCAAGACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1705  
QY 1448 ATGCTATTAATTAATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1507  
DB 1706 ATGCTATTAATTAATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1765  
QY 1508 GAACCTTAAGCAAAACAGCAAACTGATGATGATGATGATGATGATGATGATGATGATG 1567  
DB 1766 GAACCTTAAGCAAAACAGCAAACTGATGATGATGATGATGATGATGATGATGATGATG 1825  
QY 1568 ACTGCAAGATGGAGAGAGAGCTCTGAGAGGTTGAGATGATGATGATGATGATGATG 1627  
DB 1826 ACTGCAAGATGGAGAGAGAGCTCTGAGAGGTTGAGATGATGATGATGATGATGATG 1885  
QY 1628 TTGGCATGATTTGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 1687  
DB 1886 TTGGCATGATTTGTTTACATGATGATGATGATGATGATGATGATGATGATGATGATG 1945  
QY 1688 CCAGCATCAATGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1747  
DB 1946 CCAGCATCAATGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2005  
QY 1748 AGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1807  
DB 2006 AGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2065  
QY 1808 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1867  
DB 2066 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2125  
QY 1868 TGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1927  
DB 2126 TGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2185  
QY 1928 GAGGTTAACAACCTTAAGCTTAATGATGATGATGATGATGATGATGATGATGATGATG 1987



Db 2186 GAGGTTCACCACTTACGATTAACGATTAATGATGAGAAAGAGTGAACACCAATTTT 2245

Qy 1988 TTGCTCAGCTTGAGAAAGACTTCAGAAATCCACTTCTCGAAGCACTTATGATGAGG 2047

Db 2246 TTGCTCAGCTTGAGAAAGACTTCAGAAATCCACTTCTCGAAGCACTTATGATGAGG 2305

Qy 2048 TTACCAAGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2107

Db 2306 TTACCAAGCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2365

Qy 2108 TTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2167

Db 2366 TTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2425

Qy 2168 AGCGCTTGACCTTAACAGCTGTCCTCACTATTAACAATGAGGAGTGTCTGCTTC 2227

Db 2426 AGCGCTTGACCTTAACAGCTGTCCTCACTATTAACAATGAGGAGTGTCTGCTTC 2485

Qy 2228 CTGCTCCTTGACAGTACGCGCAAGCTGCTTTCTTGTGCGCAGAGTATTCAGAG 2287

Db 2486 CTGCTCCTTGACAGTACGCGCAAGCTGCTTTCTTGTGCGCAGAGTATTCAGAG 2545

Qy 2288 AGCAATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 2328

Db 2546 AGCAATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 2586

RESULT 2  
AY410584 2586 bp DNA linear GSS 12-DEC-2003  
LOCUS Pan troglodytes HCM3939 gene, VIRUAL TRANSCRIPT, partial sequence,  
AY410584  
ACCESSION AY410584  
VERSION AY410584.1 GI:39766552  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.  
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
Todd, M.A., Tanenbaum, D.M., Civeilio, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
AUTHORS 2 (bases 1 to 2586)  
REFERENCE Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
Todd, M.A., Tanenbaum, D.M., Civeilio, D.R., Lu, F., Murphy, B.,  
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Smirsky, J.J.,  
Adams, M.D. and Cargill, M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source  
1. 2586  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
<1..2586  
/locus\_tag="HCM3939"

ORIGIN  
Query Match 84.5%; Score 1967.2; DB 29; Length 2586;  
Best Local Similarity 85.0%; Pred. No. 0;  
Matches 1972; Conservative 0; Mismatches 349; Indels 0; Gaps 0;  
Qy 8 TTGGTGTGAACACAGGCAAGCACTTGAACCATGTTAAAGATCAAAAACAGGTTCTTCAG 67

Db 266 TTGGTGTGAACACAGGCAAGCACTTGAACCATGTTAAAGATCAAAAACAGGTTCTTCAG 325

Qy 68 GCATTATAGTAAAGTTTAAAGCACTTAACCACTTCCGGCTGACATCCCGTCCCAATGGGCTT 127

Db 326 NNN 385

Qy 128 TATATCAGTATCACTTGAATTAACCACTGATGAGAGGCAAGAACTCCGTTGAGCTC 187

Db 386 TATATCAGTATCACTTGAATTAACCACTGATGAGAGGCAAGAACTCCGTTGAGCTC 445

Qy 188 TTCTTTTCAACGAAAGATCTAATGAAAGTGTATGCTTTGATGGAACATATAT 247

Db 446 TTCTTTTCAACGAAAGATCTAATGAAAGTGTATGCTTTGATGGAACATATAT 505

Qy 248 TTTTACCTAAAGACTAAGCAAGCAAGGTTACTGAAGTTTAAAGACCGGAATGAG 307

Db 506 NNN 565

Qy 308 AGATGAGAGATTAAGATCACTTAACAATGAATCTTCACTTCACTTCACTTCTT 367

Db 566 AGATGAGAGATTAAGATCACTTAACAATGAATCTTCACTTCACTTCACTTCTT 625

Qy 368 TGCAATCTTAATATATATTTTCAAGAGGCTTTGAATAATCATGAATTTGCAATAATTG 427

Db 626 TNNNNNTCTTAATATATATTTTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 685

Qy 428 GACGAATTAATTAACCAATGACCAATGATATTCAGAGTCAAGGTTGATTT 487

Db 686 NNN 745

Qy 488 GGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 547

Db 746 GGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 805

Qy 548 TTAGCCATTAAGTCTTGAAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTT 607

Db 806 TTAGCCATTAAGTCTTGAAGTGAAGTGTGATTTGATTTGATTTGATTTGATTTGATTT 865

Qy 608 AGACAGAAACATTAATTTCAAGAACATTTTCAAGAACATTTTCAAGAACATTTTCAAG 667

Db 866 AGACAGAAACATTAATTTCAAGAACATTTTCAAGAACATTTTCAAGAACATTTTCAAG 925

Qy 668 TTACCAAGTATTAACAATTAAGACATTAAGACATTAAGACATTAAGACATTAAGACAT 727

Db 926 TTACCAAGTATTAACAATTAAGACATTAAGACATTAAGACATTAAGACATTAAGACAT 985

Qy 728 AGACACCTTTAAGAAAGCGAGCGGCTTGAAGTCACTTCTTAAGTCACTTCAAGAAAGC 787

Db 986 AGACACCTTTAAGAAAGCGAGCGGCTTGAAGTCACTTCTTAAGTCACTTCAAGAAAGC 1045

Qy 788 AATTAACAAGAGATGACCGACTTGAAGACAGCTGCTTGTGACGACGACCAAGAGAA 847

Db 1046 NNN 1105

Qy 848 GCGCGGCGCTTGGGGGACATGCGCAAGGCTTGCATGCTTCTGAGCTCTGCTATC 907

Db 1106 GCGCGGCGCTTGGGGGACATGCGCAAGGCTTGCATGCTTCTGAGCTCTGCTATC 1165

Qy 908 TTACAGCTTCACTGAATTAAGTGTATGATTTTAAAGTGAAGAACTTACCGCTTC 967

Db 1166 TTANNNTCTTAAGTGAATTAAGTGTATGATTTTAAAGTGAAGAACTTACCGCTTC 1225

Qy 968 ATTAAGACTTAAGTCAAGAGCAAGGAGGCTGAGAGTGAAGCACTTATGATTAATTC 1027

Db 1226 ATTAAGACTTAAGTCAAGAGCAAGGAGGCTGAGAGTGAAGCACTTATGATTAATTC 1285

Qy 1028 ATTAAGACTTAAGTCAAGAGCAAGGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1087

Db 1286 ATTAAGACTTAAGTCAAGAGCAAGGAGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1345

Qy 1088 TACTGCTCTTCAAGAGATTTTGAAGCAAGAAAGATTCACCAAGTGAAGAAATCAT 1147

Db 1346 TACTGCTCTTCTCAGGAGAAATTTTGCAGAAACAGAAAAGTTTCATCAAGTGTGAAAAACAN 1405  
 Qy 1148 TTGATTAACAATCCAAATTTGACAGATTGTCCTCAAGAAACAAGAGTGCCATTAATTA 1207  
 Db 1406 NTGATTAACAATCCAAATTTGACAGATTGTCCTCAAGAAACAAGAGTGCCATTAATTA 1465  
 Qy 1208 GTGTTAAGCCACTGATTAAGTGTGATCTATGATGCGAAGAAATTAAGACAGCC 1267  
 Db 1466 GTGTTAAGCCACTGATTAAGTGTGATCTATGATGCGAAGAAATTAAGACAGCC 1525  
 Qy 1268 ATTCAATTGATTAACAATTTTAAAGTTTACACAGCCATGCGCATGCAATGAGAAAG 1327  
 Db 1526 ATTCAATTGATTAACAATTTTAAAGTTTACACAGCCATGCGCATGCAATGAGAAAG 1585  
 Qy 1328 CAATTAATGATTAAGTGTGATGACAGAACTGAAAGCTTAAGAGTCTTAAGAGAAAG 1387  
 Db 1586 CAATTAATGATTAAGTGTGATGACAGAACTGAAAGCTTAAGAGTCTTAAGAGAAAG 1645  
 Qy 1388 TCACAGCAGACACCCAGATAGTGTCTGTCTGTTGTCAAGTAATCGAAGAGCAATATG 1447  
 Db 1646 TCACAGCAGACACCCAGATAGTGTCTGTCTGTTGTCAAGTAATCGAAGAGCAATATG 1705  
 Qy 1448 ATGCTATTAATAAATCTCTGTGTACAGATTGCCCTCAAGTCAAGTGTGTGCGCC 1507  
 Db 1706 ATGCTATTAATAAATCTCTGTGTACAGATTGCCCTCAAGTCAAGTGTGTGCGCC 1765  
 Qy 1508 GAACCTTAGGCAACAGCAAACTGTCAATGCGCATTTGCTACAAAGATTGCCCTACAGATGA 1567  
 Db 1766 GAACCTTAGGCAACAGCAAACTGTCAATGCGCATTTGCTACAAAGATTGCCCTACAGATGA 1825  
 Qy 1568 ACTGCAAGATGAGAGAGAGAGTCTGTGAGAGGTGTGACATCCCTCTTAAGTCTGTATATG 1627  
 Db 1826 ACTGCAAGATGAGAGAGAGAGTCTGTGAGAGGTGTGACATCCCTCTTAAGTCTGTATATG 1885  
 Qy 1628 TTGGCATGATTTGTTACCATGACATGACAGCTGGGCGGAGGTCAATGCGAGATTTTGTG 1687  
 Db 1886 TTGGCATGATTTGTTACCATGACATGACAGCTGGGCGGAGGTCAATGCGAGATTTTGTG 1945  
 Qy 1688 CCAGCATCAATGAAGAGATGACCCGCTGTCTCAAGCTGTCAATTTTCAAGATGAGAG 1747  
 Db 1946 CCAGCATCAATGAAGAGATGACCCGCTGTCTCAAGCTGTCAATTTTCAAGATGAGAG 2005  
 Qy 1748 AGGAGCTGTGATAGTGGCTCAAAAGTCTGCAAGAGGCTCTGAGGCTTGAATAGCT 1807  
 Db 2006 AGGAGCTGTGATAGTGGCTCAAAAGTCTGCAAGAGGCTCTGAGGCTTGAATAGCT 2065  
 Qy 1808 GCAATGATGATGATGCGCCAGCCGATCATGCTGTATCCGCGATGCGGTAGAGACGCGCAGC 1867  
 Db 2066 GCAATGATGATGATGCGCCAGCCGATCATGCTGTATCCGCGATGCGGTAGAGACGCGCAGC 2125  
 Qy 1868 TGAAAAACACTGTGTGATCAAGAGTCCACAGCTTTTGGATTTGTCTAAATTCATGGTA 1927  
 Db 2126 NNN 2185  
 Qy 1928 GAGGTTTACAACCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1987  
 Db 2186 NNN 2245  
 Qy 1988 TTGCTCATGCTGAGAGAGAGCTTCAAGATCCATCTTCTGTAACAGTTATGATGATGAGG 2047  
 Db 2246 TTGCTCATGCTGAGAGAGAGCTTCAAGATCCATCTTCTGTAACAGTTATGATGATGAGG 2305  
 Qy 2048 TTACACAGACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2107  
 Db 2306 TTACACAGACCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2365  
 Qy 2108 TTTTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2167  
 Db 2366 TTTTCTCCACACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2425  
 Qy 2168 AGGCTTGAAGCTGAGAGCTGAGAGCTGATGATGATGATGATGATGATGATGATGATG 2227  
 Db 2426 AGGCTTGAAGCTGAGAGCTGAGAGCTGATGATGATGATGATGATGATGATGATGATG 2485

Qy 2228 CTGCTCTTGGCAGTACGAGCCCAAGAGTGGCTTTCTGTTGGCAGAGATTAACAGAG 2287  
 Db 2486 CTGCTCTTGGCAGTACGAGCCCAAGAGTGGCTTTCTGTTGGCAGAGATTAACAGAG 2545  
 Qy 2288 AGCCAAATCTGTCACTGTCAAAACCGGCTTTACTACTCTTA 2328  
 Db 2546 AGCCAAATCTGTCACTGTCAAAACCGGCTTTACTACTCTTA 2586  
 RESULT 3  
 AY410585  
 LOCUS  
 DEFINITION  
 Mus musculus Hm3939 gene, VIRUAL TRANSCRIPT, partial sequence.  
 ACCESSION  
 AY410585  
 VERSION  
 AY410585.1 GI:39766553  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1 (bases 1 to 2589)  
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
 Ferritera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,  
 Adams, M.D. and Cargill, M.  
 Inferring nonneutral evolution from human-chimp-mouse orthologous  
 gene trios  
 JOURNAL  
 Science 302 (5652), 1960-1963 (2003)  
 PubMed  
 14671302  
 REFERENCE  
 2 (bases 1 to 2589)  
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,  
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,  
 Ferritera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,  
 Adams, M.D. and Cargill, M.  
 Direct Submission  
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 COMMENT  
 This sequence was made by sequencing genomic exons and ordering  
 them based on alignment  
 FEATURES  
 source  
 1..2589  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>2589  
 /locus\_tag="Hm3939"  
 ORIGIN  
 gene  
 Query Match 72.4%; Score 1684.6; DB 29; Length 2589;  
 Best Local Similarity 82.8%; Pred. No. 0;  
 Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;  
 Qy 6 CTTTGTGTGACACAGCAGCAAGCACTTAACCATGTTAAGATCAAAAGAGTCTTC 65  
 Db 267 CTTTGTGTGACACAGCAGCAAGCACTTAACCATGTTAAGATCAAAAGAGTCTTC 326  
 Qy 66 AGGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 125  
 Db 327 TGGCATCATTTGTAAGCTGAGCAGCAACCACTTCGCGCTGACCTCGCGCCACAGTGGGC 386  
 Qy 126 CTTATATCAGTATCACTTACCTTAACCACTGATGAGAGCAGAGAGCTCGCTCAGC 185  
 Db 387 CTTATATCAGTATCACTTACCTTAACCACTGATGAGAGCAGAGAGCTCGCTCAGC 446  
 Qy 186 TCTTCTTTTCAACAGAGATTAATGAGAGTGTGATGCTTTGATGAGAACATATT 245  
 Db 447 ACTGCTTCTTCAACAGAGATTAATGAGAGTGTGATGCTTTGATGAGAACATATT 506  
 Qy 246 ATTTTACCTTAAGAGATTAAGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTG 305  
 Db 507 GTTTTACCTTAAGAGATTAAGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTG 566

306 AGAGATGTGAGATTAACGATCACTTTAAACAATGAACTTCCACCTACATCAACCACTTG 365  
 567 GGAACACGTGAGATCAACCATCAACCTGACCAAGAGCTGCGCCACCTGCGCCACCTG 626  
 366 TTGGAGTTCTTAATATTAATTTTCAAGAGCTTTTGAATACTATGAATTTGGACAAT 425  
 627 CCTGAGTTCTTAATATTAATCAATCTTCAAGAGCTCTTGAATACTATGAATTTGGACAAT 686  
 426 TGAAGAAATTAATTAACCAATGAGCCAAATGATTAATCAAGTCAAGTGTGAT 485  
 687 TGAAGAAATTAATTAACCAATGAGCCAAATGATTAATCAAGTCAAGTGTGAT 746  
 486 TTGGCTGCTTCACTACTTCCATCTTCAATGAGAAACGATCATGCTCTGACATGA 545  
 747 CTGGCCCGCTTCAACCTCATCTTCAATGAGAAACGATCATGCTCTGACATGA 806  
 546 CGTTAGCAATTAAGCTTGAAGAGAGCTGTTTGAATTCATGTTCACTTTATCA 605  
 807 CGTCAAGCAAGAGTCTCCGACGAGACTCTCTAATCTTCAATGTTCAATCTAACA 866  
 606 TCAGACAGAAAGCAATTAATTTCAAGAACAGTTTCCAAAGACTAATAGGTTTATG 665  
 867 GCAAGCAGAGAGCAAGTTTCAAGAGCAAGTTCAGAGAGCTCATAGGCTCATGCT 926  
 666 TCTTACCAATTAACATTAAGACATACAGAGTGAATTAATGATCTGGGACCAAGATCC 725  
 927 TCTCAACCAAGTCAATTAACAGACCTACCGGATGATGATGATGATGATGATGATG 986  
 726 CAAGAGCACTTTAAGAAAGCGAGCTGAAATCACTTTTGAATTAATCAAGAA 785  
 987 AAGAGCACTTTAAGAAAGCGAGCTGAAATCACTTTTGAATTAATCAAGAA 1046  
 786 GCAATTAACAACAGAGATCACCACTTGAAGAGCTCTTGGTCAAGCCCAAG 845  
 1047 GCAATTAACAACAGAGATCACCACTTGAAGAGCTCTTGGTCAAGCCCAAG 1106  
 846 AAGAGCGAGCGCTGAGGAGCACTGCAAGGCTGCAATGCTTCTGAGCTCTGTA 905  
 1107 GAGGAGAGCGCGCGGCGGCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1166  
 906 TCTTACAGCTTACTGATTAATATCGTATGATTTTAACTGATGAAGAATTAAGCT 965  
 1167 TCTTACAGCTTACTGATTAATATCGTATGATTTTAACTGATGAAGAATTAAGCT 1226  
 966 TCAATCAAGCTTACTGATTAATATCGTATGATTTTAACTGATGAAGAATTAAGCT 1025  
 1227 GCAACAGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1286  
 1026 TCAATTAACAACAGAGATCACCACTTGAAGAGCTCTTGGTCAAGCCCAAG 1085  
 1287 GCAATTAACAACAGAGATCACCACTTGAAGAGCTCTTGGTCAAGCCCAAG 1346  
 1086 CTTTACTGCTCTTCAAGAGAAATTTGGAACAAGAAAGATTCACCAAGTGAAGAAAC 1145  
 1347 CTTTACTGCTCTTCAAGAGAAATTTGGAACAAGAAAGATTCACCAAGTGAAGAAAC 1406  
 1146 ATTGATTAACAATTCACCAATTTGAGATTTGTCACCAAGAAACAGAGTGCACATTAAT 1205  
 1407 GTTGTATTAACAACCAATTTGAGATTTGTCACCAAGAAACAGAGTGCACATTAAT 1466  
 1206 TATGTTAAGCACTAGATTAATGCTGTTGATCTATAGCGAAGAAATTAAGAGAGC 1265  
 1467 CAGCGTGAAGCACTAGATTAATGCTGTTGATCTATAGCGAAGAAATTAAGAGAGC 1526  
 1266 CAATTAATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1325  
 1527 CAATTAATTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1586  
 1326 AGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1385  
 1587 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1646

1386 GGTCAACAGCAGACACCAGATAGTTGTCTGTCTGTTTCAAGTAAATCCGAAGACAAATA 1445  
 1647 GGTGAGTCAAGACACCTCAATAGTTGTCTGTCTGTTTCAAGTAAATCCGAAGACAAATA 1706  
 1446 CGATGCTATTAATAAATACTGTGTACAGATAGCCCTACCCCAAGTCAATGTGTGTG 1505  
 1707 TGAATGCAATCAAGAGTAACTGTGTACAGATAGCCCTACCCCAAGTCAATGTGTG 1766  
 1506 CCGAATCTTAAGCAACAGCAACTGTGTACAGATAGCCCTACCCCAAGTCAATGTGTG 1565  
 1767 CCGAATCTTAAGCAACAGCAACTGTGTACAGATAGCCCTACCCCAAGTCAATGTGTG 1826  
 1566 GAACTGCAAGATGAGAGAGAGTGTGAGAGGTGACATCCCTGCAAGTGTGTGTAT 1625  
 1827 GAACTGCAAGATGAGAGAGAGTGTGAGAGGTGACATCCCTGCAAGTGTGTGTAT 1886  
 1626 CGTGGCATGATGATTTGATCAATGACATGACAGCTGGGCGAGAGTCAATGCAAGATTTGT 1685  
 1887 CGTGGCATGATGATTTGATCAATGACATGACAGCTGGGCGAGAGTCAATGCAAGATTTGT 1946  
 1686 TGGCAAGATCAATGAGAGATGACCGGCTGATTCAGCTGATTAATTAAGATTAAGAG 1745  
 1947 CGCAGATCAATGAGAGATGACCGGCTGATTCAGCTGATTAATTAAGATTAAGAG 2006  
 1746 ACAAGAGCTGTGATGAGGCTCAAGTGTGCTGCAAGGCTCTGAGGCTTTGAATAG 1805  
 2007 GCAAGAGCTGTGATGAGGCTCAAGTGTGCTGCAAGGCTCTGAGGCTTTGAATAG 2066  
 1806 CTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1865  
 2067 CTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2126  
 1866 GCTGAATAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1925  
 2127 GCTGAATAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2186  
 1926 TAAAGTTACACCTTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 1985  
 2187 GAAAGTTACACCTTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2246  
 1986 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2045  
 2247 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2106  
 2046 GGTTAACACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2165  
 2207 AGTCAACACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2226  
 2106 TGTTCCTCCCAACATTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2285  
 2267 TGTTCCTCCCAACATTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2346  
 2166 ACAAGCTTACCTTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2225  
 2427 CCAAGCTTACCTTAAGATTAAGGATTAATGATGATGATGATGATGATGATGATGATG 2286  
 2226 TCTGCTCTTCCCAAGATTAAGGATTAATGATGATGATGATGATGATGATGATGATG 2345  
 2487 CCGTCAACCTTCCCAAGATTAAGGATTAATGATGATGATGATGATGATGATGATGATG 2406  
 2286 AGAGCCAAATCTGCTCAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2345  
 2547 AGAGCCAAATCTGCTCAAGATTAAGGATTAATGATGATGATGATGATGATGATGATG 2406

RESULT 4  
 BF66315 874 bp mRNA linear EST 21-DEC-2000  
 LOCUS 602119233F1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4276136 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF66315  
 VERSION BF66315.1 GI:11940210  
 KEYWORDS EST.



```

Db      2  CAAGTGGAAAAAATTGATTACATCCACAAATTGCAAGTGGTCAAAAAGAAAACAGA 61
QY      1192  GGTGACCACTTAAATAGTGTAGAGCCACTAGATACTGGCTGTGATCTATACCGAGA 1251
Db      62  GGTGACCACTTAAATAGTGTAGAGCCACTAGATACTGGCTGTGATCTATACCGAGA 121
QY      1252  AATTATGAAGCCCAATTCATTGATCAAAATCTATTATTAAGTTACACCGAGCCATGGGC 1311
Db      122  AATTATGAAGCCCAATTCATTGATCAAAATCTATTATTAAGTTACACCGAGCCATGGGC 181
QY      1312  ATGCAATATGAGAAAAGCAATATGATGATGATGATGATGATGATGATGATGATGATGAT 1371
Db      182  ATGCAATATGAGAAAAGCAATATGATGATGATGATGATGATGATGATGATGATGATGAT 241
QY      1372  GTCTTACAGCAAAAGGTCAACAGCAGACACCCAGATGATGATGATGATGATGATGATGAT 1431
Db      242  GTCTTACAGCAAAAGGTCAACAGCAGATCCAGATGATGATGATGATGATGATGATGATGAT 301
QY      1432  CGGAGAGCAAAATAGATGCTATTAAATAATACCTGTGATGATGATGATGATGATGATGAT 1491
Db      302  CGAAGAGCAAAATAGATGCTATTAAATAATCTATGATGATGATGATGATGATGATGATGAT 361
QY      1492  CAGTGTGTGTGTGCCCCGCAACCTTAGGCAAAAGCAAGCACTGTGATGATGATGATGAT 1551
Db      362  CAGTGTGTGTGTGCCCCGCAACCTTAGGCAAAAGCAAGCACTGTGATGATGATGATGAT 421
QY      1552  ATTGGCTTACAGATTAATCTGCAAGATGGAGAGAGAGCTCTGAGAGGTGGAATCCCTCG 1611
Db      422  ATGCGCTTACAGATTAATCTGCAAGATGGAGAGAGAGCTCTGAGAGGTGGAATCCCTCG 481
QY      1612  AAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1671
Db      482  AAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 541
QY      1672  ATGCGAGATTTGTTGTCGACGATCATATGAGGATG 1707
Db      542  ATGCGAGATTTGTTGTCGACGATCATATGAGGATG 577

RESULT 6
LOCUS   BM751055 527 bp mRNA linear EST 04-MAR-2002
DEFINITION K-BST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5',
ACCESSION BM751055
VERSION   BM751055.1 GI:19080673
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS  Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE    21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT  Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boseun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: B column: 08
High quality sequence stop: 527.
Location/Qualifiers
1..527
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S9SNU601-21-E08"

```

```

/sex="M"
/tissue_type="neutrophils"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/clone_id="S9SNU601"
/note="Organ: Stomach; Vector: pME18-FU3; Site 1: XhoI;
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deacapped
with tobacco acid pyrophosphatase (TAP). The deacapped
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII-digested pME18-FU3 vector. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

```

```

ORIGIN
Query Match 22.1%; Score 513.4; DB 12; Length 527;
Best Local Similarity 99.6%; Pred. No. 5.2e-135;
Matches 525; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY      930  GCGTAATGATTTTAAAGTATGAAAGACTTAGCGTTCTATACAGACTACTACGAGCA 989
Db      1  GCGTAATGATTTTAAAGTATGAAAGACTTAGCGTTCTATACAGACTACTACGAGCA 60
QY      990  AAGCAGCGTGAAGTGGAGCACTCATTTGATTAATCAATTAATAAAGATATGTTCAAG 1049
Db      61  AAGCAGCGTGAAGTGGAGCACTCATTTGATTAATCAATTAATAAAGATATGTTCAAG 120
QY      1050  GAGCTTCCAGATGAGGCTTTGAGCTTTGATTTCCAACTACTGCTCTCCAGAGGAAT 1109
Db      121  GAGCTTCCAGATGAGGCTTTGAGCTTTGATTTCCAACTACTGCTCTCCAGAGGAAT 180
QY      1110  TTTCAGAAAGAGAAAGATTCACCAAGTGGAGAAACATTTGATTAATCAATTCATTTGC 1169
Db      181  TTTCAGAAAGAGAAAGATTCACCAAGTGGAGAAACATTTGATTAATCAATTCATTTGC 240
QY      1170  AGATTGGTCCAAAGAAAGAGAGTGCACCTTAATTAATGTTAAAGCCATGATTAACG 1229
Db      241  AGATTGGTCCAAAGAAAGAGAGTGCACCTTAATTAATGTTAAAGCCATGATTAACG 300
QY      1230  GCTGTGATCTATACGCAAGAAATATGAGAGCCAAATTCATGATCAAAATCTATT 1289
Db      301  GCTGTGATCTATACGCAAGAAATATGAGAGCCAAATTCATGATCAAAATCTATT 360
QY      1290  TAAAGTTACACGACCATGGGCAATGCAATAGAGAAAGCAATATGATGATGATGATG 1349
Db      361  TAAAGTTACACGACCATGGGCAATGCAATAGAGAAAGCAATATGATGATGATGATG 420
QY      1350  CAGACTGAGAGCTTACTTAAGAGTCTTAAGAGCAAAAGGTCAACGAGACCCAGATAG 1409
Db      421  CAGACTGAGAGCTTACTTAAGAGTCTTAAGAGCAAAAGGTCAACGAGACCCAGATAG 480
QY      1410  TGTCTGTCTGTTGTCAAGTAATC-GGAAGAGCAATATGATGATGATGATGATGATGAT 1455
Db      481  TGTCTGTCTGTTGTCAAGTAATC-GGAAGAGCAATATGATGATGATGATGATGATGAT 527

RESULT 7
LOCUS   BF664941 1495 bp mRNA linear EST 21-DEC-2000
DEFINITION 602119191F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5',
ACCESSION BF664941
VERSION   BF664941.1 GI:11938836
KEYWORDS EST.
SOURCE   Homo sapiens (human)

```



## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

1 (bases 1 to 1495)

## AUTHORS

NIH-MGC http://mgs.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LNCM1096 row: F column: 24

High quality sequence stop: 511.

## FEATURES

## source

1..1495

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4276247"

/tissue\_type="Primitive neuroectoderm"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_1lb="NIH\_MGC\_56"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site\_1:

SfiI (ggcgccctcgcc); Site\_2: SfiI (ggccattatggc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 21.6%; Score 502.4; DB 10; Length 1495;

Best Local Similarity 94.7%; Pred. No. 1.3e-131;

Matches 553; Conservative 0; Mismatches 26; Indels 5; Gaps 3;

## FEATURES

## source

1..705

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4075561"

/tissue\_type="Glioblastoma"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_1lb="NIH\_MGC\_57"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site\_1:

SfiI (ggcgccctcgcc); Site\_2: SfiI (ggccattatggc);

Double-stranded cDNA was prepared from cell line RNA. 5'

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGCCGACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

## ORIGIN

Query Match 21.1%; Score 492; DB 10; Length 705;

Best Local Similarity 93.8%; Pred. No. 8e-149;

Matches 621; Conservative 0; Mismatches 30; Indels 11; Gaps 10;

QY 1333 ATGATTGAAGTGTGATGACAGAACTGAACTTAAAGTCTTACAGAAAGGTGCA 1392

DB 68 GAGAGACCCAGATAGTGTCTGTCTGTGTCAGATATCGAAGAGCAAAATACGATGT 127

QY 1453 ATTAAATAATCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCGAAAC 1512

DB 128 ATTAATAAATCTGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCGAAAC 187

QY 1513 TTAGGCAAAACAGAACTGTCTGTCTGTGTACAAAGATTGCCCTTACAGATGATGC 1572

DB 188 TTAGGCAAAACAGAACTGTCTGTCTGTGTACAAAGATTGCCCTTACAGATGATGC 247

QY 1573 AAGATGGAGAGAGCTCTGAGAGGTGAGACATCCCTGAGAGCTGTATGATGTTGGC 1632

DB 248 AAGATGGAGAGAGCTCTGAGAGGTGAGACATCCCTGAGAGCTGTATGATGTTGGC 307

QY 1633 ATGATTTTACATGACATGACAGCTGGGCGAGAGTCAATGCGAGATTGTTGCCAGC 1692

DB 308 ATGATTTTACATGACATGACAGCTGGGCGAGAGTCAATGCGAGATTGTTGCCAGC 367

QY 1693 ATGATGAAGAGATGACCGCGTGTCTCAGCGTGCATATTTACAGATAGAGACAGAG 1752

DB 368 ATGATGAAGAGATGACCGCGTGTCTCAGCGTGCATATTTACAGATAGAGACAGAG 427

QY 1753 CTGTGATGATGGCTCAAAAGTCTGCTGCAAGGCGCTTGAAGGCTTGGAATAGTGCAT 1812

QY 1333 ATGATTGAAGTGTGATGACAGAACTGAACTTAAAGTCTTACAGAAAGGTGCA 1392

DB 11 AGGATTGAAGTGTGATGACAGAACTGAACTTAAAGTCTTACAGAAAGGTGCA 70

QY 1393 GCAGACACCAAGATAGTGTCTGTCTGTGTCAAGTATCGAAGAGCAAAATACGATGT 1452

DB 71 GCAGACACCAAGATAGTGTCTGTCTGTGTCAAGTATCGAAGAGCAAAATACGATGT 129



QY 1453 ATTAAATAATACCTGTGTACAGATTGCCCTAACCCCAAGTCACTGTGTGTGCGCCGAAC 1512  
 DB 130 ATTTAAAAATACCTGTGTACAGATTGCCCTAACCCCAAGTCACTGTGTGTGCGCCGAAC 189  
 QY 1513 TTAGGCAAAACAGCAAACTGTCTCATGGCCATTGTCTCAAAAGATTGCCCTAACAGTAACTGC 1572  
 DB 190 TTAGGCAAAACAGCAAACTGTCTCATGGCCATTGTCTCAAAAGATTGCCCTAACAGTAACTGC 249  
 QY 1572 AAGATGGAGAGAGAGCTGTGAGGGGTGACATCCCGCTGAAGCTGTGTGTGTGTGTC 1632  
 DB 250 AAGATGGAGAGAGAGCTGTGAGGGGTGACATCCCGCTGAAGCTGTGTGTGTGTGTC 308  
 QY 1633 ATCGATTGTTCACATGACATGACAGCT-GGGCGAGGTTCATGCGAGGA-TTTGTGTGCCA 1690  
 DB 309 ATCGATTGTTCACATGACATGACAGCTGGGGCGGAGGTTCATGCGAGGATTTTGTGCCA 368  
 QY 1691 GGATCAATGAAGGATGACCCGCTGTCTCAGCTGCAATTTTCAAGATGAGAGACAG 1750  
 DB 369 GGATCAATGAAGGATGACCCGCT-GTCTCAGCTGCAATTTTCAAGATGAGAGACAG 427  
 QY 1751 AGCTGTGATGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCT-TGGAAATAGCT-- 1807  
 DB 428 AGCTGTGATGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTGTGGAATAGCTTG 487  
 QY 1808 GCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1866  
 DB 488 CAATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547  
 QY 1867 CTG-AAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1924  
 DB 548 TTGAAAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607  
 QY 1925 GTAGAGGTTTACACCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1984  
 DB 608 GTAGAGGTTTACACCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667  
 QY 1985 TT 1986  
 DB 668 TT 669

RESULT 9  
 BF080923 533 bp mRNA linear EST 18-OCT-2000  
 LOCUS 233690 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BF080923  
 VERSION BF080923.1 GI:10874744  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 533)  
 Fahnensterg, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,  
 Vallet, J., Wise, T., Rohrer, G.A., Perrea, G., Sultana, R.,  
 Quackenbush, J., and Keefe, J.W.  
 Porcine gene discovery by normalized cDNA-library sequencing and  
 EST cluster assembly  
 Mamm. Genome 13 (8), 475-478 (2002)  
 JOURNAL Mamm. Genome 13 (8), 475-478 (2002)  
 MEDLINE 22213789  
 PUBLISHED 12226715  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAAAGAGTATGACACG  
 BACKWARD: GTTCCACATGACGACG

FEATURES  
 Source  
 1. 533  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /issue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_id="MARC 2P1G"  
 /note="Vector: pCMV SPORTE; Site 1: NotI; Site 2: SalI;  
 library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."

ORIGIN  
 Query Match 20.2%; Score 471.2; DB 10; Length 533;  
 Best Local Similarity 92.9%; Pred. No. 6.1e-123; Indels 0; Gaps 0;  
 Matches 494; Conservative 0; Mismatches 38;

QY 1132 CAAGGTGAAAAAATTGATTAACAATCCAAATTTGACAGATTGTCGCAAGAAACAGAGA 1191  
 DB 2 CAAGGTGAAAAAATTGATTAACAATCCAAATTTGACAGATTGTCGCAAGAAACAGAGA 61  
 QY 1192 GGTGACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1251  
 DB 62 GGTGACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 QY 1252 AATTATGACAGCCCAATTCATGATTAACAATCTATTTAAAGTTACACAGCCATGAGC 1311  
 DB 122 AATTATGACAGCCCAATTCATGATTAACAATCTATTTAAAGTTACACAGCCATGAGC 181  
 QY 1312 ATGCAATGAGAAAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1371  
 DB 182 ATGCAATGAGAAAGCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 QY 1372 GTCTTACAGAAAAGTCAAGACAGACAGACCCAGATGTTGTCTGTCTGTGTCAAGTAAT 1431  
 DB 242 GTCTTACAGAAAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301  
 QY 1432 CGAAGAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1491  
 DB 302 CGAAGAGCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361  
 QY 1492 CAGGTGTGTGTGTCGCGAAGCTTACAGCAACAGCAAGCTGATGAGCATTTGCTACAAAG 1551  
 DB 362 CAGGTGTGTGTGTCGCGAAGCTTACAGCAACAGCAAGCTGATGAGCATTTGCTACAAAG 421  
 QY 1552 ATTGCCCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611  
 DB 422 ATTGCCCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481  
 QY 1612 AAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1663  
 DB 482 AAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 533

RESULT 10  
 AK030116 3559 bp mRNA linear HTC 18-SEP-2003  
 LOCUS AK030116  
 DEFINITION Mus musculus adult male testis cDNA, Riken full-length enriched  
 library, clone:4932443D1 product:piwi like homolog 1  
 (Drosophila)-like, full insert sequence.  
 ACCESSION AK030116  
 VERSION AK030116.1 GI:26326094  
 KEYWORDS HTC; CAP trapper  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)





## ORIGIN

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

Query Match 19.0%; Score 441.4; DB 9; Length 462;  
Best Local Similarity 99.6%; Pred. No. 1.9e-114;  
Matches 453; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

1714 TGGTTCACGCTGCATATTTCAAGATAGAGACAGAGCTGTAGATGGCTCAAAATC 1773  
1 TGGTTCACGCTGCATATTTCAAGATAGAGACAGAGCTGTAGATGGCTCAAAATC 60  
1774 TGCCTGACAGCGCTCTGAGGCTTGAATAGCTGCAATGATACATGCCAGCCGATC 1833  
61 TGCCTGACAGCGCTCTGAGGCTTGAATAGCTGCAATGATACATGCCAGCCGATC 120  
1834 ATGCTGACCGCATGCGCTGAGAGAGCGGCACTGAAAACATGCTGAATAGAGAGT 1893  
121 ATGCTGACCGCATGCGCTGAGAGAGCGGCACTGAAAACATGCTGAATAGAGAGT 179  
1894 CCACAGTTTTGATGCTAAATCCATGTTAGAGGTTACAACTTACCTAGACTAGCGTA 1953  
180 CCACAGTTTTGATGCTAAATCCATGTTAGAGGTTACAACTTACCTAGACTAGCGTA 239  
1954 ATTGCTGAGAAAAGAGTGAACACCAATTTTGTCTGATCTGAGAGAACTTCAG 2013  
240 ATTGCTGAGAAAAGAGTGAACACCAATTTTGTCTGATCTGAGAGAACTTCAG 299  
2014 AATCAGTCTCTGAGACGTTATGATGATGATGATGATGATGATGATGATGATGAT 2073  
300 AATCAGTCTCTGAGACGTTATGATGATGATGATGATGATGATGATGATGATGAT 359  
2074 TTTATCGTGAAGCAGGCTGTGAGAGTGTGTTCTTCCACATTCATTAATGTCATC 2133  
360 TTTATCGTGAAGCAGGCTGTGAGAGTGTGTTCTTCCACATTCATTAATGTCATC 419  
2134 TATGACAAACAGCGGCTGAGAGCGAGACCAATACA 2168  
420 TATGACAAACAGCGGCTGAGAGCGAGACCAATACA 454

RESULT 12  
LOCUS Bg773137 818 bp mRNA linear EST 15-MAY-2001  
DEFINITION 6072721862P1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4838918 5',  
mRNA sequence.  
ACCESSION Bg773137  
VERSION Bg773137.1 GI:14083790  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 818)  
TITLE NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgsaps-remail.nih.gov  
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM10773 row: k column: 15

FEATURES  
source

High quality sequence stop: 789.  
Location/Qualifiers  
1..818

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4838918"  
/lab\_host="DH10B"  
/clone\_id="NIH\_MGC\_97"  
/note="Organ: testis; Vector: pBluescriptDR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gagag); Oligo-dT primed using primer 5'-TTTTTTTTTTT-3', size-selected for average insert size 2.2 kb and normalized for R07 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 18.8%; Score 438.8; DB 12; Length 818;  
Best Local Similarity 97.3%; Pred. No. 1.4e-113;  
Matches 500; Conservative 0; Mismatches 7; Indels 7; Gaps 5;

8 TTGCTGACACCAAGGACGACCTTACACCTGTTAAAGATCAAAAACAGGTTCTCAG 67  
305 TTGCTGACACCAAGGACGACCTTACACCTGTTAAAGATCAAAAACAGGTTCTCAG 364  
68 GCATTATGTAAGGTTAAGACCTAACCTTCCGCTGACATCCCGTCCAGTGGGCT 127  
365 GCATTATGTAAGGTTAAGACCTAACCTTCCGCTGACATCCCGTCCAGTGGGCT 424  
128 TATATCACTATCACTGATTAACCACTGATGAGAGCCAGAAAGCTCCGTTCAAGCTC 187  
425 TATATCACTATCACTGATTAACCACTGATGAGAGCCAGAAAGCTCCGTTCAAGCTC 484  
188 TTTCTTTTCAACAGAAATGTAATGTAAGAGTCACTTTTGTATGAGACGATATTAT 247  
485 TTTCTTTTCAACAGAAATGTAATGTAAGAGTCACTTTTGTATGAGACGATATTAT 544  
248 TTTTACTTAAAGACTACAGCAAAAGGTTACTGAAATTTTATGTAAGCCCGGAATGAG 307  
545 TTTTACTTAAAGACTACAGCAAAAGGTTACTGAAATTTTATGTAAGCCCGGAATGAG 604  
308 AGGATGTAGATTAACGATCACTTTTACAA-TGAACTTCCACCTACATCACTGAT 366  
605 AGGATGTAGATTAACGATCACTTTTACAA-TGAACTTCCACCTACATCACTGAT 664  
367 TTGCAATCTATATATATTTTTCAGAGGC-TTTTGAATATCATGAATTTGC---ACA 422  
665 TTGCAATCTATATATATTTTTCAGAGGC-TTTTGAATATCATGAATTTGC---ACA 724  
423 AATTGACGAATTTTATTAACCCAAATGACCCAA-TTGATTTTCAAGTCAAGGTTG 481  
725 AATTGACGAATTTTATTAACCCAAATGACCCAA-TTGATTTTCAAGTCAAGGTTG 784  
482 TGAATTTGG-CCTGGCTTCACTACTTCATCCCTTC 514  
785 TGAATTTGGCTTGGCTTCACTACTTCATCCCTTC 818

RESULT 13  
LOCUS BU854142 642 bp mRNA linear EST 16-OCT-2002  
DEFINITION AGENCOURT 10402932 NIH\_MGC\_82 Homo sapiens cDNA clone IMAGE:6621314  
5', mRNA sequence.  
ACCESSION BU854142  
VERSION BU854142.1 GI:24039108  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



QY 1797 TTGAAATAGCTGCAATGAGTACATGCGCCAGGATCATGCTGACCGGATGCGGATGAG 1856  
 DB 121 TTGAGAGTGGCTGCATGAAATACATGCGCCAGGATCATGCTGACCGGATGCGGATGAG 180  
 QY 1857 AGACGGCCGAGCTGAAAACCTGCTGAACTACGAAAGTGCACAGTTTGGATTGCTAAA 1916  
 DB 181 GAGCGGAGCTGAAAGCCTGCTGAACTACGAAAGTGCACAGTTTGGATTGCTAAA 240  
 QY 1917 ATCCATTTGTAAGAGTTTACCAACCTGAGTAACTGAGTAACTGAGTAACTGAGTAA 1976  
 DB 241 GTGAGTGGGAGAGTTTACCAACCTGAGTAACTGAGTAACTGAGTAACTGAGTAA 300  
 QY 1977 CACCAATTTTGGCTGCTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 2036  
 DB 301 TGCCAGGATTTTGGCTGCTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 360  
 QY 2037 TGAATGAGAGTTTACCAACCTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 2096  
 DB 361 CGATGAGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 420  
 QY 2097 AAGTGTAGAGTTTCTCCACACTTACATGCTGAGTAACTGAGTAACTGAGTAACTGAGTAA 2156  
 DB 421 AAGGAGAGTGTGCTCCACACTTACATGCTGAGTAACTGAGTAACTGAGTAACTGAGTAA 480  
 QY 2157 AGACCATATACAGGCTTGAACCTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 2216  
 DB 481 CGACCATATACAGGCTTGAACCTGAGTAACTGAGTAACTGAGTAACTGAGTAACTGAGTAA 540  
 QY 2217 CATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2240  
 DB 541 CATGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564

RESULT 15  
 AL041473/c 413 bp mRNA linear EST 04-SEP-2003  
 LOCUS DKFZP434E2417.81 434 (synonym: htes3) Homo sapiens cDNA clone  
 DEFINITION DKFZP434E2417.3, mRNA sequence.  
 AL041473  
 ACCESSION AL041473.1 GI:5420824  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 413)  
 Bloecher, H., Boecher, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and  
 Wiemann, S.  
 EST (Bloecher, et al.)  
 TITLE EST  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: MIPS  
 MIPS  
 Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany  
 This is the 3' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; Ltd.,  
 Braunschweig/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.  
 r1 sequence also available.  
 This clone (DKFZP434E2417) is available at the RZPD in Berlin.  
 Please contact the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..413  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZP434E2417"  
 /tissue\_type="testis"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_id="434 (synonym: htes3)"

ORIGIN /note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"  
 Query Match 17.7%; Score 413; DB 9; Length 413;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-106;  
 Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1637 ATTGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1696  
 DB 413 ATTGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354  
 QY 1697 ATGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1756  
 DB 353 ATGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294  
 QY 1757 TAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1816  
 DB 293 TAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234  
 QY 1817 ACATGCGCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1876  
 DB 233 ACATGCGCAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 174  
 QY 1877 TGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1936  
 DB 173 TGGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 114  
 QY 1937 ACCCTGAGCTAAGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1996  
 DB 113 ACCCTGAGCTAAGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 54  
 QY 1997 CTGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2049  
 DB 53 CTGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

Search completed: May 8, 2004, 15:48:52  
 Job time: 5991 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: May 5, 2004, 15:06:00 ; Search time 60 seconds  
(without alignments)  
3649.572 Million cell updates/sec

Title: US-10-043-774B-2  
Perfect score: 4102  
Sequence: 1 MIFGNTNRQNDHYKESKMG.....VGQSHREPNLSLNRLYYL 775

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: Geneseq\_29Jan04: \*  
2: Geneseqp1980s: \*  
3: Geneseqp1980s: \*  
4: Geneseqp2000s: \*  
5: Geneseqp2000s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4102	100.0	775	ADB94987	ADB94987 Human hae
2	4061	99.0	861	AAV90235	AAV90235 Human p1w
3	3857	96.5	862	AAV90234	AAV90234 Mouse p1w
4	3883	94.7	829	ADB65285	ADB65285 Human pro
5	2764	67.4	523	ABP67995	ABP67995 Human col
6	2764	67.4	523	ABU61933	ABU61933 Human Hiw
7	2101.5	51.2	852	ADB65492	ADB65492 Human pro
8	1468	35.8	866	ABP62102	ABP62102 Drosophila
9	1468	35.8	866	ABU61934	ABU61934 Drosophila
10	1434	35.0	844	ABU61937	ABU61937 Fruitfly
11	1434	35.0	844	ADB94988	ADB94988 Fruitfly
12	1428.5	34.8	843	ABP62084	ABP62084 Drosophila
13	1416.5	34.5	843	AAV90233	AAV90233 Drosophila
14	1389.5	33.9	580	AAU07866	AAU07866 Polypepti
15	1324.5	33.3	498	ADP43865	ADP43865 Human mRN
16	1324.5	33.3	498	ABU61938	ABU61938 Human ful
17	1295.5	31.6	824	ABU61936	ABU61936 C. elegan
18	1230	30.0	530	ABP94209	ABP94209 Human pro
19	1162	28.3	722	ABU61935	ABU61935 C. elegan
20	962	22.5	371	AAV92722	AAV92722 Human pro
21	802	15.6	642	ADA54973	ADA54973 Human pro
22	652	15.9	130	AAV58977	AAV58977 Breast an
23	580	14.1	341	ABG05957	ABG05957 Novel hum
24	515.5	12.6	915	ADB17496	ADB17496 Soybean p
25	510	12.4	868	AAV42119	AAV42119 Arabidops

25	510	12.4	1048	3	AAV42118	AAV42118 Arabidops
27	510	12.4	1048	6	ADB17512	ADB17512 A. thalia
28	508.5	12.4	859	5	ABP64718	ABP64718 Human pro
29	508.5	12.4	860	5	ABG97473	ABG97473 Human NKA
30	507.5	12.3	1101	6	ADB17510	ADB17510 Rice post
31	506.5	12.3	857	7	ADB81572	ADB81572 Human euk
32	506	12.3	186	7	ABU61932	ABU61932 Human par
33	503	12.3	884	3	AAV52594	AAV52594 Arabidops
34	503	12.3	906	3	AAV51563	AAV51563 Arabidops
35	503	12.3	983	3	AAV51562	AAV51562 Arabidops
36	503	12.3	988	3	AAV51561	AAV51561 Arabidops
37	503	12.3	1064	3	AAV52593	AAV52593 Arabidops
38	500	12.2	950	4	ABP67046	ABP67046 Drosophila
39	500	12.2	984	4	ABP62573	ABP62573 Drosophila
40	500	12.2	984	4	ABP67045	ABP67045 Drosophila
41	500	12.2	984	5	ABP70016	ABP70016 Larval vi
42	498.5	12.2	1058	6	ADB17494	ADB17494 Soybean p
43	497	12.1	816	6	AAV57044	AAV57044 Human nuc
44	497	12.1	860	4	AAV60292	AAV60292 Human pro
45	497	12.1	860	4	AAV94291	AAV94291 Human pro

## ALIGNMENTS

RESULT 1  
ADB94987  
ID ADB94987 standard; protein; 775 AA.  
XX  
AC ADB94987;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human haematopoietic growth regulatory protein, HIWI.  
XX  
KW Human, haematopoietic growth regulatory protein; HIWI; cytostatic;  
KW stem cell therapy; bone marrow; peripheral blood;  
KW CD34+ haematopoietic stem cell; leukaemia; cancer.  
XX  
OS Homo sapiens.  
XX  
PN US2003129599-A1.  
XX  
PD 10-JUL-2003.  
XX  
PF 10-JAN-2002; 2002US-00043774.  
XX  
PR 10-JAN-2002; 2002US-00043774.  
XX  
PA (SHAR/) SHARMA A.  
PA (HOFF/) HOFFMAN R.  
XX  
PI Sharma A, Hoffman R,  
XX  
DR WPI; 2003-744346/70.  
DR N-PSDB; ADB94986.  
XX  
PT New hiwi nucleic acids, for screening compounds which modulate hiwi,  
PT useful as cancer therapeutic.  
XX  
PS Claim 2; Fig 1; 39pp; English.  
XX  
CC The invention relates to a nucleic acid comprising a nucleotide sequence  
CC encoding a human hiwi protein (a haematopoietic growth regulatory gene).  
CC Also included are the hiwi protein, the hiwi cDNA (used as a probe), a  
CC recombinant expression construct comprising the hiwi cDNA, a cell culture  
CC transformed with the recombinant expression construct, screening a  
CC compound for modulating human hiwi gene activity in cells expressing the  
CC human hiwi gene product, a cell membrane preparation (or cytosolic  
CC preparation or nuclear preparation) comprising a human hiwi gene product  
CC (or derivative) and identifying a compound that induces or increases hiwi  
CC gene expression in mammalian cells. Culturing bone marrow or peripheral  
CC blood culture in the presence of compounds identified by the methods is

CC useful for increasing retention of primitive CD34+ haematopoietic stem  
 CC cells. Alternatively the method comprises introducing a retroviral or  
 CC lentiviral expression construct into the cells. The compounds are useful  
 CC for haematopoietic stem cell development and maintenance and for  
 CC inhibiting leukemia cell growth and treating cancer. The present  
 CC sequence represents the hwi protein.

XX  
 XX  
 SQ Sequence 775 AA;

Query Match 100.0%; Score 4102; DB 7; Length 775;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIFGNTNRLNDHVKESKTSSTGIIYRLSTNHFRLTSRPMALYOHIDNPLMEARRR 60  
 DB 1 MIFGNTNRLNDHVKESKTSSTGIIYRLSTNHFRLTSRPMALYOHIDNPLMEARRR 60  
 QY 61 SALLFQHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRIITLTNELPPTSP 120  
 DB 61 SALLFQHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRIITLTNELPPTSP 120  
 QY 121 TCIQFYNIIFRRLIKIMNLQOIGRNYNPNDDIDPSHRUVIWPFGTTSILOEYSIMLC 180  
 DB 121 TCIQFYNIIFRRLIKIMNLQOIGRNYNPNDDIDPSHRUVIWPFGTTSILOEYSIMLC 180  
 QY 181 TDVSHKVLRSFTVLDPMENFYHOTESEKFOQVSKELIGLWLTKNKTYRVDDIDMDQ 240  
 DB 181 TDVSHKVLRSFTVLDPMENFYHOTESEKFOQVSKELIGLWLTKNKTYRVDDIDMDQ 240  
 QY 241 NPKSTFKKADGSEVSFLEYRKYNOEITDLKOPVLVSOPRRRPGGTLPGPAMILPEL 300  
 DB 241 NPKSTFKKADGSEVSFLEYRKYNOEITDLKOPVLVSOPRRRPGGTLPGPAMILPEL 300  
 QY 301 CYLIGLIDPKANDNNWMDLAVHTRLTPBOQREVERGLIYHKNNDVVOBELMDGLSPD 360  
 DB 301 CYLIGLIDPKANDNNWMDLAVHTRLTPBOQREVERGLIYHKNNDVVOBELMDGLSPD 360  
 QY 361 SNLSFSGRILOTEKIHQSGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420  
 DB 361 SNLSFSGRILOTEKIHQSGKTFDYNPQFADMSKETRGAPLISYKPLDNMLIYTRNRYE 420  
 QY 421 AANSLIONLKYTPAMKQKRAKAIIEVDRTKAYLVLQOKTADQIIVCLISNRKD 480  
 DB 421 AANSLIONLKYTPAMKQKRAKAIIEVDRTKAYLVLQOKTADQIIVCLISNRKD 480  
 QY 481 KYDAIKKYLCTDCPTPSCVAVARTLGGQOTVMAATKIALQMNCKMGSELNRVDIPKLV 540  
 DB 481 KYDAIKKYLCTDCPTPSCVAVARTLGGQOTVMAATKIALQMNCKMGSELNRVDIPKLV 540  
 QY 541 MIVGIDCYHMTAGRSIAGFVASINSGMTWFSRCLIFODNGOELVNDGLKYLQALRAW 600  
 DB 541 MIVGIDCYHMTAGRSIAGFVASINSGMTWFSRCLIFODNGOELVNDGLKYLQALRAW 600  
 QY 601 NSCNVMSRSRIIVVDGVDGQDKTLVYVEVPQFADCLKSIGRGYNPLFIYVKKENV 660  
 DB 601 NSCNVMSRSRIIVVDGVDGQDKTLVYVEVPQFADCLKSIGRGYNPLFIYVKKENV 660  
 QY 661 RFFAOSGRLQNPFLPGTVIDVEVTRPEWYDFEIVSQAVRSGSVSTHNVIVDMSGKPD 720  
 DB 661 RFFAOSGRLQNPFLPGTVIDVEVTRPEWYDFEIVSQAVRSGSVSTHNVIVDMSGKPD 720  
 QY 721 HIGRLTYLCHITYYMPPGIVRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775  
 DB 721 HIGRLTYLCHITYYMPPGIVRPACQYAHKLAFLVGSIHREPLSLSNRLTYL 775

RESULT 2

AA90235 standard; protein; 861 AA.

AC AA90235;  
 XX 29-AUG-2000 (first entry)  
 DT

XX Human piwi protein, designated hwi.  
 DE  
 XX piwi family protein; piwi; miwi; hwi; gene therapy; tissue dystrophy;  
 KW anaemia; immunodeficiency; male infertility; human.  
 XX  
 XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 76

FT /label= 'Leu, Ile

FT /note= "encoded by NTA"

FT Misc-difference 303

FT /label= 'Leu, Ile

FT /note= "encoded by NTA"

FT Misc-difference 735

FT /label= 'Leu, Ile

FT /note= "encoded by NTA"

XX WO20032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US028764.

XX 04-DEC-1998; 98US-0110901P.

XX (UYDU-) UNIV DUKE.

XX Lin H;

XX WPI; 2000-412085/35.

XX N-PSDB; AAA07588.

XX piwi family nucleic acids, polypeptides, and antibodies, useful in gene

XX therapy of diseases such as cancer and in various research and diagnostic

XX applications.

XX claim 4; Page 189-194; 201pp; English.

XX This sequence represents the human piwi family protein, designated hwi.

XX The piwi family nucleic acids and polypeptides are used in gene therapy

XX of diseases such as cancer and also in various research and diagnostic

XX applications. The sequences can also be used to treat tissue dystrophy,

XX anaemia, immunodeficiency, and male infertility

XX SQ Sequence 861 AA;

Query Match 99.0%; Score 4061; DB 3; Length 861;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 768; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GYNTNRLNDHVKESKTSSTGIIYRLSTNHFRLTSRPMALYOHIDNPLMEARRRSL 63  
 DB 90 GYNTNRLNDHVKESKTSSTGIIYRLSTNHFRLTSRPMALYOHIDNPLMEARRRSL 149  
 QY 64 LFOHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRIITLTNELPPTSPCL 123  
 DB 150 LFOHEDLIGKCHAFDGTILFLPKLQOKTEVFSKTRNGEDVRIITLTNELPPTSPCL 209  
 QY 124 QFYNIIFRRLIKIMNLQOIGRNYNPNDDIDPSHRUVIWPFGTTSILOEYSIMLCYV 183  
 DB 210 QFYNIIFRRLIKIMNLQOIGRNYNPNDDIDPSHRUVIWPFGTTSILOEYSIMLCYV 269  
 QY 184 SHKVLSEFTVLDPMENFYHOTESEKFOQVSKELIGLWLTKNKTYRVDDIDMDQNK 243  
 DB 270 SHKVLSEFTVLDPMENFYHOTESEKFOQVSKELIGLWLTKNKTYRVDDIDMDQNK 329  
 QY 244 STFKADGSEVSFLEYRKYNOEITDLKOPVLVSOPRRRPGGTLPGPAMILPELCL 303  
 DB 330 STFKADGSEVSFLEYRKYNOEITDLKOPVLVSOPRRRPGGTLPGPAMILPELCL 389  
 QY 304 TGLIDPKANDNNWMDLAVHTRLTPBOQREVERGLIYHKNNDVVOBELMDGLSPD 363

DB 390 TGLTDMKNDNFVMDLAVHTRLTPEOROREVGLIDYTHKNDVQRELRDMSLSDSL 449  
 QY 364 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 423  
 DB 450 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 509  
 QY 424 SLIQLFKVTPAMGOMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 483  
 DB 510 SLIQLFKVTPAMGOMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 569  
 QY 484 AIKKYLCCTDPTSCQCVARTLGKQQTVAATKIALQNNCKMGELMRVDIPLKLVMIY 543  
 DB 570 AIKKYLCCTDPTSCQCVARTLGKQQTVAATKIALQNNCKMGELMRVDIPLKLVMIY 629  
 QY 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRQGLVDGLKVCLOALRAMNSC 603  
 DB 630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRQGLVDGLKVCLOALRAMNSC 689  
 QY 604 NEYPSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRFF 663  
 DB 690 NEYPSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRFF 749  
 QY 664 AOSGRLQNPPLPGTVIDVETREPEWDFPIVSAVRSQVSPTHYNYIYDNGSLKPDHIQ 723  
 DB 750 AOSGRLQNPPLPGTVIDVETREPEWDFPIVSAVRSQVSPTHYNYIYDNGSLKPDHIQ 809  
 QY 724 RLTYKLCCHYYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 775  
 DB 810 RLTYKLCCHYYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 861

RESULT 3  
 AAY90234  
 ID AAY90234 standard; protein; 862 AA.  
 XX AAY90234;  
 XX 29-AUG-2000 (first entry)  
 XX Mouse piwi protein, designated miwi.  
 XX  
 XX piwi family protein; miwi; miwi; gene therapy; tissue dystrophy;  
 XX anaemia; immunodeficiency; male infertility; mouse; ds.  
 XX  
 XX Mus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 90 /label= Leu, Ile  
 FT /note= "encoded by NTT"  
 FT Misc-difference 216 /note= "unspecified amino acid; encoded by NTC"  
 FT Misc-difference 383 /label= Leu, Ile  
 FT /note= "encoded by NTC"  
 FT Misc-difference 816 /label= Leu, Ile  
 FT /note= "encoded by NTC"  
 FT /note= "encoded by NTC"  
 XX  
 XX MO200032039-AI.  
 XX  
 XX 08-JUN-2000.  
 XX  
 XX 03-DEC-1999; 99WO-US028764.  
 XX  
 XX 04-DEC-1998; 98US-0110901P.  
 XX  
 XX (UYDU-) UNIV DUKE.  
 XX  
 XX Lin H;  
 XX  
 XX WPI; 2000-412085/35.  
 XX  
 XX

DR N-PSDB; AAA07587.  
 XX  
 XX piwi family nucleic acids, polypeptides, and antibodies, useful in gene  
 PT therapy of diseases such as cancer and in various research and diagnostic  
 PT applications.  
 XX  
 XX  
 XX  
 PS Claim 4; Page 180-185; 201pp; English.  
 XX  
 XX This sequence represents the mouse piwi family protein, designated miwi.  
 CC The piwi family nucleic acids and polypeptides are used in gene therapy  
 CC of diseases such as cancer and also in various research and diagnostic  
 CC applications. The sequences can also be used to treat tissue dystrophy,  
 CC anaemia, immunodeficiency, and male infertility  
 XX  
 XX Sequence 862 AA;  
 SQ  
 Query Match 95.5%; Score 3957; DB 3; Length 862;  
 Best Local Similarity 95.9%; Pred. No. 0;  
 Matches 740; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

4 GYNTRONLDHYVESKTSGGIIVRLSTNHFRITSPQALYQYHIDVNPIMBARLBSAL 63  
 91 GYNTRONLDHYVESKTSGGIIVKLSNHFRITSPQALYQYHIDVNPIMBARLBSAL 150  
 64 LFQEDDLTGKCHAFPGTILFLPKLQOKYVEFSKTRGEVDRIITLTNLEPPTSPTCL 123  
 151 LFQEDDLTGKCHAFPGTILFLPKLQOKYVEFSKTRGEVDRIITLTNLEPPTSPTCL 210  
 124 GFYNIIFRLIKIMLQOIGRYNYNPNDIPDSRLVWPGFTTSLIYENSIMLCTDV 183  
 211 GFYNIIFRLIKIMLQOIGRYNYNPNDIPDSRLVWPGFTTSLIYENSIMLCTDV 270  
 184 SHKVLRSFTVLDPMENFTHQTEHKKPOEVSKEELIGLYVLTKNKTYRVDDIDWDQNK 243  
 271 SHKVLRSFTVLDPMENFTHQTEHKKPOEVSKEELIGLYVLTKNKTYRVDDIDWDQNK 330  
 244 STEKKADGSEVSEFLFYRKQYQCEITDLPQVLPQSPKRRPGGTLPEPAMLIPELCYL 303  
 331 STEKKADGSEVSEFLFYRKQYQCEITDLPQVLPQSPKRRPGGTLPEPAMLIPELCYL 390  
 304 TGLTDMKNDNFVMDLAVHTRLTPEOROREVGLIDYTHKNDVQRELRDMSLSDSL 363  
 391 TGLTDMKNDNFVMDLAVHTRLTPEOROREVGLIDYTHKNDVQRELRDMSLSDSL 450  
 364 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 423  
 451 LSPSGRILOTEKIHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNLIIYTRRYEAA 510  
 424 SLIQLFKVTPAMGOMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 483  
 511 SLIQLFKVTPAMGOMRKAIMIEVDRTAAYLRVLOQKTAOTQIVCLLSNRDOKD 570  
 484 AIKKYLCCTDPTSCQCVARTLGKQQTVAATKIALQNNCKMGELMRVDIPLKLVMIY 543  
 571 AIKKYLCCTDPTSCQCVARTLGKQQTVAATKIALQNNCKMGELMRVDIPLKLVMIY 630  
 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRQGLVDGLKVCLOALRAMNSC 603  
 631 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRQGLVDGLKVCLOALRAMNSC 690  
 604 NEYPSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRFF 663  
 691 NEYPSRIIVYRDGVGDGQKTLVNYEVPQFLDCKSIGGYNPRLTVIVKKRVNTRFF 750  
 664 AOSGRLQNPPLPGTVIDVETREPEWDFPIVSAVRSQVSPTHYNYIYDNGSLKPDHIQ 723  
 751 AOSGRLQNPPLPGTVIDVETREPEWDFPIVSAVRSQVSPTHYNYIYDNGSLKPDHIQ 810  
 724 RLTYKLCCHYYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 775  
 811 RLTYKLCCHYYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLSNRLYYL 862

RESULT 4  
 ADB65285  
 ID ADB65285 standard; protein; 829 AA.  
 XX  
 AC ADB65285;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human protein encoded by clone TEST120060830.  
 XX  
 KM Human, pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KM cell regeneration; membrane protein; signal transduction-related protein;  
 KM transcription-related protein; osteoporosis; neurological disease;  
 KM cancer; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI308459-A2.  
 XX  
 PD 07-MAY-2003.  
 XX  
 PF 28-MAR-2002; 2002EP-00007401.  
 XX  
 PR 05-NOV-2001; 2001JP-00379298.  
 PR 25-JAN-2002; 2002US-00350378.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX  
 PI Isogai T, Sugiyama T, Otsuki T, Makamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Negahari K, Masuno Y;  
 XX  
 DR WPI; 2003-450961/43.  
 DR N-PSDB; ADB63315.  
 XX  
 PT New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX  
 PS Claim 1; Page; 222pp; English.  
 XX  
 CC The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease transduction-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX  
 SQ Sequence 829 AA;

Query Match 94.7%; Score 3883; DB 7; Length 829;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 733; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	GVNTRONLDHYKESKTSGGIIIVRLSTNHPSLTSPQWALVQYHIDVNFLEARRLSAL	63
DB	90	GVNTRONLDHYKESKTSGGIIIVRLSTNHPSLTSPQWALVQYHIDVNFLEARRLSAL	149
QY	64	LPQHEDLIGKCHAPDGTLLPLPKRLQOKVTEVFSKTRNGEDVRITITVNTNLPPTSPTCL	123
DB	150	LPQHEDLIGKCHAPDGTLLPLPKRLQOKVTEVFSKTRNGEDVRITITVNTNLPPTSPTCL	209
QY	124	QFYNIIFPRLLKINMLQOIGRNNYNNPDIDIPSHRLVYMPGFPTSILQYNSIMLCTDV	183
DB	210	QFYNIIFPRLLKINMLQOIGRNNYNNPDIDIPSHRLVYMPGFPTSILQYNSIMLCTDV	269
QY	184	SHKVLRSSTVLDFMNFYHQTSEHKFQSVSKELIGLVITKYNKTYRVDIDMDQNPX	243
DB	270	SHKVLRSSTVLDFMNFYHQTSEHKFQSVSKELIGLVITKYNKTYRVDIDMDQNPX	329
QY	244	STFKKADSEVSELEYAKQYNOETITDKQVLYSOPRRRPGGTLPGPAMLIPELCYL	303
DB	330	STFKKADSEVSELEYAKQYNOETITDKQVLYSOPRRRPGGTLPGPAMLIPELCYL	389
QY	304	TGLTDKXENDFNVMKDLAVHTRLTPEQFQREVGRLIDYIHKNQVQRELRDWSJSPDNL	363
DB	390	TGLTDKXENDFNVMKDLAVHTRLTPEQFQREVGRLIDYIHKNQVQRELRDWSJSPDNL	449
QY	364	LSFSGRLLQTEKIHQGGKTFDYNQPADWSKETGAPLISKPLDNNLLITRENYEAN	423
DB	450	LSFSGRLLQTEKIHQGGKTFDYNQPADWSKETGAPLISKPLDNNLLITRENYEAN	509
QY	424	SLIQNLFVYTPAMGOMKAKAIMIEVDRETEAVLYVLQOKTADTIQIVVCLLSNRKQYD	483
DB	510	SLIQNLFVYTPAMGOMKAKAIMIEVDRETEAVLYVLQOKTADTIQIVVCLLSNRKQYD	569
QY	484	AIKXLLCTDCTPSCQVARTLQKQTYMAIATIALQANCKMGEELRVDPILKLVITV	543
DB	570	AIKXLLCTDCTPSCQVARTLQKQTYMAIATIALQANCKMGEELRVDPILKLVITV	629
QY	544	GIDCHDMTAAGRSIAAGVASINSGMTRMFCIFQDQGLVYGLKVCLOALRAANSC	603
DB	630	GIDCHDMTAAGRSIAAGVASINSGMTRMFCIFQDQGLVYGLKVCLOALRAANSC	689
QY	604	NEVMSRIIVRDGVGDQLTLVNVYVPOFLDLKSIIGRINPRLTYIVKKNVTRFE	663
DB	690	NEVMSRIIVRDGVGDQLTLVNVYVPOFLDLKSIIGRINPRLTYIVKKNVTRFE	749
QY	664	AQSGRLQNPUGTVIVDEYTRPEMYPETVSAVRSGSVPTKNNVYDNSGKPDHIQ	723
DB	750	AQSGRLQNPUGTVIVDEYTRPEMYPETVSAVRSGSVPTKNNVYDNSGKPDHIQ	809
QY	724	RLTYVLCIIYNNMP	737
DB	810	RLTYVLCIIYNNMP	823

RESULT 5  
 ADB67995  
 ID ADB67995 standard; protein; 523 AA.  
 XX  
 AC ADB67995;  
 XX  
 DT 13-DEC-2002 (first entry)  
 XX  
 DE Human colon cancer related polypeptide SEQ ID NO 2603.  
 XX  
 KM Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200258534-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 16-NOV-2001; 2001WO-US043704.

20-NOV-2000; 2000US-0252222P.  
 06-FEB-2001; 2001US-0267011P.  
 28-MAR-2001; 2001US-0279670P.  
 10-JUL-2001; 2001US-0304037P.  
 (CORI-) CORIXA CORP.  
 Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GB;  
 WPI: 2002-608400/65.  
 N-PDB; ABV89287.  
 New isolated tumor colon polynucleotide and polypeptide, useful for the  
 diagnosis, prevention and/or treatment of cancer, in particular colon  
 cancer.  
 Claim 2; SEQ ID NO 2603; 266pp + Sequence listing; English.  
 The invention relates to a human colon tumor expressed polynucleotide  
 (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of  
 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)  
 complements of (i); (iii) at least 20 contiguous residues of (i); (iv)  
 sequences that hybridize to (i), under moderately stringent conditions;  
 (v) sequences having at least 75% or 90% identity to (i); or (vi)  
 degenerate variants of (i). The compositions and methods of the present  
 invention are useful for the diagnosis, prevention and/or treatment of  
 cancer, particularly colon cancer. (i) can be used in gene therapy and  
 (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.  
 Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WFO  
 at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 523 AA;  
 Query Match 67.4%; Score 2764; DB 5; Length 523;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-260;  
 Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

254 VSFLEYRKNQNOETDLDKQPVLVQPKRRGPGGTLDPAPMLPELCYLTGLTDKMD 313  
 2 VSFLEYRKNQNOETDLDKQPVLVQPKRRGPGGTLDPAPMLPELCYLTGLTDKMD 61  
 314 FVMKDLAVHTLTTEOROREVGRLLIDYHKNDYORLEBDMGSPDSNLSFSGRIQT 373  
 62 FVMKDLAVHTLTTEOROREVGRLLIDYHKNDYORLEBDMGSPDSNLSFSGRIQT 121  
 374 EKHOGGKTFDYNPQFADMSKETRGAPLISVPLDNMLLIYTRRYEANSILQNTFRYT 433  
 122 EKHOGGKTFDYNPQFADMSKETRGAPLISVPLDNMLLIYTRRYEANSILQNTFRYT 181  
 434 PAMGQMRKAIMIEVDRTREATLRLQOKTADTQIVVCLSSNKKDYDAIKKYLCDG 493  
 182 PAMGQMRKAIMIEVDRTREATLRLQOKTADTQIVVCLSSNKKDYDAIKKYLCDG 241  
 494 PPSOCVAVARTLGKQOTWAIATKIALQNCMGSELWVDPILKLVMIVGIDCYHDMTA 553  
 242 PPSOCVAVARTLGKQOTWAIATKIALQNCMGSELWVDPILKLVMIVGIDCYHDMTA 301  
 554 GRRSIAGFVASTINEGMTWFSRCIFQDRGQELVDGLKVCLOALPAMNSCNEIMPRIIV 613  
 302 GRRSIAGFVASTINEGMTWFSRCIFQDRGQELVDGLKVCLOALPAMNSCNEIMPRIIV 361  
 614 YADGVDGDKLTLYNVEVPQFADCKSIGRGNPRLTYIVVKGKRVNTFFPAGSGRLONP 673  
 362 YADGVDGDKLTLYNVEVPQFADCKSIGRGNPRLTYIVVKGKRVNTFFPAGSGRLONP 421  
 674 LFGTVIDVEVTRPEVYDFTVSQAIVSSGVSPTHTNVIVYNSGLKPHDIQRLTYLCHIY 733  
 422 LFGTVIDVEVTRPEVYDFTVSQAIVSSGVSPTHTNVIVYNSGLKPHDIQRLTYLCHIY 481  
 734 YNMPGVIRVPACQYAHKLAFLVGSIHREPLSTSNLTYL 775  
 482 YNMPGVIRVPACQYAHKLAFLVGSIHREPLSTSNLTYL 523

RESULT 6  
 ID ABU61933 standard; protein; 523 AA.  
 AC ABU61933;  
 DT 19-AUG-2003 (first entry)  
 DE Human Hwi protein.  
 XX Human; Hwi-Hy; stem cell maintenance factor; immunosuppressive;  
 XX anti-diabetic; antiparkinsonian; cerebroprotective; neuroprotective;  
 XX diabetes; CNS disease; graft-versus-host disease; Parkinson's disease;  
 XX organ transplantation; biological-based sensor; Hwi.  
 OS Homo sapiens.  
 XX US2003027255-A1.  
 XX 06-FEB-2003.  
 XX 26-OCT-2001; 2001US-00029495.  
 XX 20-AUG-1999; 99US-00378667.  
 XX (BALI/) BALINGER D G.  
 XX (ORMA/) DRMANAC R T.  
 XX (LABA/) LABAT I.  
 XX (STAC/) STACHE-CRAIN B.  
 XX (DICK/) DICKSON M C.  
 XX (JONE/) JONES L W.  
 XX (XUEA/) XUE A.  
 XX (TANG/) TANG Y T.  
 XX (LIUC/) LIU C.  
 XX (ASUN/) ASUNDI V.  
 XX Ballinger DG, Drmanac RT, Labat I, Stache-Crain B, Dickson MC;  
 XX Jones LW, Xue A, Tang YT, Liu C, Asundi V;  
 XX WPI: 2003-479536/45.  
 XX New isolated stem cell maintenance factor polynucleotide and polypeptide,  
 XX useful for the diagnosis and treatment of disorders such as diabetes, CNS  
 XX diseases, graft versus host disease, Parkinson's disease and organ  
 XX transplantation.  
 XX Disclosure; Fig 1; 63pp; English.  
 XX The invention relates to a new isolated polynucleotide comprising: (a) the  
 XX human partial Hwi-Hy cDNA appearing as AC62541; (b) having the stem  
 XX cell maintenance factor protein coding nucleotide sequence of a  
 XX polynucleotide of (a); or (c) having the mature stem cell maintenance  
 XX factor protein coding nucleotide sequence of (a). Also include are an  
 XX isolated polynucleotide encoding a polypeptide with stem cell maintenance  
 XX factor activity that hybridizes under stringent conditions to the  
 XX complement of (or has a sequence identity greater than 90% to AC62541).  
 XX an expression vector comprising the Hwi-Hy cDNA, a host cell genetically  
 XX engineered to contain or express the Hwi-Hy cDNA in operative  
 XX association with a regulatory sequence that controls expression of the  
 XX polynucleotide in the host cell, the partial Hwi-Hy stem cell  
 XX maintenance factor appearing as ABU61932, an anti-Hwi-Hy antibody,  
 XX detecting Hwi-Hy in a sample (comprising contacting the sample with a  
 XX compound that binds to and form a complex with the polypeptide),  
 XX identifying the binding compound and modulating cell survival and/or  
 XX death (comprising contacting the cell with Hwi-Hy for a time sufficient  
 XX to modulate the activity). The methods and compositions of the present  
 XX invention are useful for the diagnosis and treatment of disorders such as  
 XX diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease  
 XX and organ transplantation. They can also be used for re-engineering  
 XX damaged or diseased tissues, manufacture of pharmaceuticals and in the  
 XX development of biological-based sensors. The present sequence is a





QY 418 NTEAANSLQNLFKVTPAMQMRKAIMIEVDRTAYLRLVQKVTADTQIVCLISN 477  
 DB 495 TEYVAESFNLCLRRVAGSGFVNDYPKIKIKVQENPAFAFRAIQVYVDPDVLVWCLIPSN 554  
 QY 478 RDKXDAIKKYLCTCPSPSCVAVRTLGKQOTVNAIKTALQNKCKMGEIWRVDP 537  
 DB 555 QRTYDSIKKILSSDCEVPSQCLARKTKQGMNLSIKTKAMQMTCKLGGELVAVEIPL 614  
 QY 538 KLVMTVGIDCYDMTAGRSISNGFVASINEGMTWFSRCIFQDRQGLVDGLKYCLQAL 597  
 DB 615 KSLMTVGIDVCKDALSKQVMVGVASVNPRTTRFSCICIQRTMTVDADCLKFMGAL 674  
 QY 598 RAMNSCNEMBSRIIVYRDVGDGQKTLVNTVEVQPLDCLKSIGRGVNPFLTVYKCR 657  
 DB 675 NKWYKKNHDLPRRIIVYRAGVGDGQLETLIEYVQPLLSVAESGNSNTSRLSVYVRKK 734  
 QY 658 VTRFFAQSGLQNPFGTVIDEVTRPEWYDFEIVSQAVRSGSVSPTHNVVYDNGSL 717  
 DB 735 CMRFETEMKRTVQNPPLGTVDSATRENTDFILISQVACRGTVSEFTYNNVYDDNGL 794  
 QY 718 KPDHIOQLTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLNELYYL 775  
 DB 795 KPDHIOQLTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLNELYYL 852

## RESULT 8

ABB62102  
 ID ABB62102 standard; protein; 866 AA.

XX ABB62102;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 13098.

KM Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

PE 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

FR 11-JUL-2000; 2000US-00614150.

PA (PERE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW,

DR WPI; 2001-6556860/75.

XX N-PSDB; ABL06205.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX interactions.

PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutic and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB16511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-

CC AB577072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences

SQ Sequence 866 AA;

Query Match 35.8%; Score 1468; DB 4; Length 866;

Best Local Similarity 39.4%; Pred. No. 2e-133;

Matches 305; Conservative 154; Mismatches 291; Indels 24; Gaps 12;

QY 17 SKTSSGSIIVRLSTNHFRLTRSPOMALYQYHIDYNELMEARLSALFQHEDLIGKCHA 76

DB 102 SKGVGVGTHITVQANFKYAKRNMTIYQRYVDFTDVAETRLRRFLEHKHLLGG-YI 160

QY 77 FQGTILFLP--KRLQOK--VTEVFSKTRNGEDVIRITITLNEPPTSPTCLQFNIIER 131

DB 161 FQGTINFCINQFAVQDPSVYELVTKSRAGENIEIKIRAVGSQSTDAEQFVNLILIR 220

QY 132 RLKIKMLQOIGANNYNPNPDIDIPSHRLVIMGFETSILOVENSIMLCDSVSHKYLASE 191

DB 221 RAMEGDLKLVSYTYDDPOAKINLENFRQQLNFPGYTISRQENDILLSEICHKVMRTB 280

QY 192 TVLDFMNFYHQTEHKEFOEVSKEILGLVLTIKYNNKTYRVDIDMDQNPXSTFKKADG 251

DB 281 TLVNIISDAIRSD--YQSTFKRAVMGVILTDYNNKTYRIDVDVFOGTPLCKFKTNDG 338

QY 252 SEVSFELEYRKQYNOEITDKQPVLSQPKRRRGPGTLPGAMLIPELCYLTGLTKDKR 311

DB 339 -EISYDYTKKRNITIRDLKQPLVMSRPTDKVIRGN--DQALMITIPELARATGMIDNR 396

QY 312 NDFNVKDLAVHTRLTPEQROREVGRLIDYIHKNDVQRELRDWGLSPDSNLISFSGRI 371

DB 397 ADERTLRAMSEHTRLPDRRIERLMFNRLKSCQSVETLSKWNIELDSALVEIPARVL 456

QY 372 QTEKIHQGGKTPDYNQPADNSKETRGAPLISVXPLDMTLLLYTRANYAANSLQNLK 431

DB 457 PPEKILFNGQKIFVCDARADMTNEBRTCSMFKNVINRWYVITPSNLTRETOEFVQMCIR 516

QY 432 VTPAMQMRKAIMIEV--DRTAAYLRVQKVTADTQIVCLISNRKQXDAIKKYIC 490

DB 517 TASSMGNICNPIYEBIPDRNGTYSQALDNAAADPOLVWVYMSFNEKISCIKKRC 576

QY 491 TDCPTPSQCVARLTGKQGT---VMAIATKIALQNKCKMGEIWRVDPKLVMIVGID 546

DB 577 VDRPVPQVYTLKVLAIPRQCKPTGLMSIATKVIQMAKMGAPQVAFPLHGLMTVGSD 636

QY 547 CYHDMTAGRSIAGFVASINBGMT--RMPFSRCIFQDRQGLVDGLKYCLQALRAMNSCBE 605

DB 637 VCHSPKMKKSYGAVVATNDQKESFRFYSVNEHITKGQELSQMSVNAACALRSYQBOQR 696

QY 606 YMPSRIVYRDVGDGQKTLVNTVEVQPLDCL---KSIGRGVNPFLTVYKRVNTR 661

DB 697 SLPERILFPRDGVGGQLYQVNVSEVNTLKDRLDIYSAGKQEGCRMTFIIVSKRINSR 756

QY 662 FFAQSGGLQNPFGTVIDEVTRPEWYDFEIVSQAVRSGSVSPTHNVVYDNGSLKDPH 721

DB 757 YF--TGRH--NPPGVVDDVITLPERYDFPLVSAQVIRIGVSPSYNISDMGMLADNK 812

QY 722 IQRLLTYKCHLYNMPGVIRVAPCOYAHKLAFLVGSIHREPNLSLNELYYL 775

DB 813 LQMLSYKQTHMYNNSGTRIVPAVCHYAHKLAFLVASINRAPSAGLQNGQLYFL 866

RESULT 9  
 ABB61934  
 ID ABB61934 standard; protein; 866 AA.

XX ABB61934;

AC 19-AUG-2003 (first entry)

XX Fruitfly Aubergerine protein.

XX Hiw1-HY; stem cell maintenance factor; immunosuppressive; antidiabetic;

KW antiparkinsonian; cerebroprotective; neuroprotective; diabetes;

KW CNS disease; graft-versus-host disease; Parkinson's disease;

KW organ transplantation; biological-based sensor; fruitfly; aubergerine.



PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONES) JONES L W.  
PA (XUEA) XUE A T.  
PA (TANG/) TANG Y T.  
PA (LIUC) LIU C.  
PA (ASUND) ASUNDI V.  
XX  
XX  
PI Ballinger DG, Drmanac RT, Labat I, Stache-Crain B, Dickson MC,  
PI Jones LW, Xue A, Tang YT, Liu C, Asundi V;  
XX  
XX WPI; 2003-479556/45.  
DR  
PT New isolated stem cell maintenance factor polynucleotide and polypeptide,  
PT useful for the diagnosis and treatment of disorders such as diabetes, CNS  
PT diseases, graft versus host disease, Parkinson's disease and organ  
PT transplantation.  
XX  
PS Disclosure; Fig 1; 63pp; English.  
XX  
XX The invention relates to a new isolated polynucleotide comprises: (a) the  
CC human partial Hiw1-Hy cDNA appearing as AC662541; (b) having the stem  
CC cell maintenance factor protein coding nucleotide sequence of a  
CC polynucleotide of (a); or (c) having the mature stem cell maintenance  
CC factor protein coding nucleotide sequence of (a). Also include are an  
CC isolated polynucleotide encoding a polypeptide with stem cell maintenance  
CC factor activity that hybridizes under stringent conditions to the  
CC complement of (or has a sequence identity greater than 90% to AC662541),  
CC an expression vector comprising the Hiw1-Hy cDNA, a host cell genetically  
CC engineered to contain or express the Hiw1-Hy cDNA in operative  
CC association with a regulatory sequence that controls expression of the  
CC polynucleotide in the host cell, the partial Hiw1-Hy stem cell  
CC maintenance factor appearing as AB061932, an anti-Hiwi-Hy antibody,  
CC detecting Hiwi-Hy in a sample (comprising contacting the sample with a  
CC compound that binds to and form a complex with the polypeptide),  
CC identifying the binding compound and modulating cell survival and/or  
CC death (comprising contacting the cell with Hiw1-Hy for a time sufficient  
CC to modulate the activity). The methods and compositions of the present  
CC invention are useful for the diagnosis and treatment of disorders such as  
CC diabetes, CNS diseases, graft-versus-host disease, Parkinson's disease  
CC and organ transplantation. They can also be used for re-engineering  
CC damaged or diseased tissues, manufacture of pharmaceuticals and in the  
CC development of biological-based sensors. The present sequence is a  
CC homologous sequence to human Hiw1-Hy  
XX  
XX  
SQ Sequence 844 AA;  
Query Match 35.0%; Score 1434; DB 7; Length 844;  
Best Local Similarity 38.0%; Pred. No. 4e-130;  
Matches 294; Conservative 162; Mismatches 266; Indels 32; Gaps 13  
QY 17 SKTGGSGGIYRSLSTNHFRLTSPQALYQYHIDYVPLMEARRRLRSALLPQHEDLTKCHA 76  
DB 88 SKKGTGVPVPMQATNFRLKTRPEMRIYVHAYEPEPSTLENFRVAGVLSNHALLSGTL 147  
QY 77 FDCGTLFLPKLQOKTEVEFSKTRNGEDVRLITLTMELPEPTSCQAFNIIFRRLKI 136  
DB 148 FDDGLQFTTRKPEQELTVLVSXKSIDIYKISIKRVGISCABRPFLQVNLITRRSMKG 207  
QY 137 MATQIGRNVYAPNDPIDPSRLRYVYNGPTTSIIQYNSIMCTDSVSHVYASEVLDP 196  
DB 208 LIMEIVGRNLPEDFRALIEIRFERMELWPGVETSIKQHEKDLILGTELTIRHVMTEITLID 267  
QY 197 MEFNYQTEHKEHQEOVSKELIGLVLTTKYNNKTYRVDDIDWDQNPKSTPKADGSEVSF 256  
DB 268 MRCSSNHPARH--QDEVRVNVVLDLITLDYNNKRYRIDVDFPGQTPKSTP-SCCKGDISF 324  
QY 257 LEYVRQYQNEITDLKQPLVYQOPKRRBPGSTLGPAMLPELCVYTGTLGDKGRDPNV 316  
DB 325 VEYLTKNIRRIDHNPILISK-NEDKRLKTNASSELVYVLPCLCRVTGTAAMKRSNFDL 383  
QY 317 MCKLAIVTRLTPBORQREVGRLLIDYIKNDVQ--RELKDWGLSFDNSLISFGRILQT 373

[illegible]

CC human h1wi gene product, a cell membrane preparation (or cytosolic  
CC preparation or nuclear preparation) comprising a human h1wi gene product  
CC (or derivative) and identifying a compound that induces or increases h1wi  
CC gene expression in mammalian cells. Culturing bone marrow or peripheral  
CC blood culture in the presence of compounds identified by the methods is  
CC useful for increasing retention of primitive CD34+ haematopoietic stem  
CC cells. Alternatively the method comprises introducing a retroviral or  
CC lentiviral expression construct into the cells. The compounds are useful  
CC for haematopoietic stem cell development and maintenance and for  
CC inhibiting leukaemia cell growth and treating cancer. The present  
CC sequence represents the fruitfly p1wi protein, a homologue of the H1wi  
CC protein.

XX Sequence 844 AA;

XX Query Match 35.0%; Score 1434; DB 7; Length 844;

XX Best Local Similarity 38.0%; Pred. No. 4e-130;

XX Matches 294; Conservative 162; Mismatches 286; Indels 32; Gaps 13;

QY 17 SKTSSGIIYRLSTNHFRLTSRPMALYOYHIDYPLMEARLRSLFQHEDLIGKCHA 76  
DB 88 SKKGTGVPMVLMQTNFRLKTPKEMRIYHVFESPIENPRKGVLSNANLIGSGL 147  
QY 77 FDGILFLPKLQOKVTEVSKTRNGEDVRIITITLNLPTSPCLQFYNIIFRLIKI 136  
DB 148 FDGLQFLTRKFEQETIVLSGSKLDIEYKISIKFVGFSICAPRFLQVNLILRSKMG 207  
QY 137 KNLQOIGRNYNPNDDIPSRHLVYPGFTSIILOYENSIMLCITVSHKVLRSFVDF 196  
DB 208 LNLVLGRNLPDPRAKIEIRPEKMEIMPGETSIRKHEKIDILGTETIKVMTETIYI 267  
QY 197 MENFYQTEBEHKQEOVSKELIGVYLTKNKNTYRVDDIDWQNKSTFKKADGSEVSF 256  
DB 268 MRCCHNPARH--QDEVKRVNVDLIVITDNNRTYRINDVDFQGTFRKSTF--SCKGRDISF 324  
QY 257 LEYRKQYNOEITDLKQPVIVSOPKRRRGGTLPKPMILPILCYITGILDMKRDENV 316  
DB 325 VEYLTKNYRIRIRHNPPLISK-NRDKALKTNASELVAVIPELCRVGTGLANMERNFOL 383  
QY 317 MKDLAVTRLTPEBOROREVRLDIYHKNNVQ---RELADWGLSPSNLISFSGRILQ 373  
DB 384 MRAVSTYRNNPKOR---TDRLRAFNRLQTPESVAVLDMWELDXVTEVQSKIIQ 440  
QY 374 EKI--HOGGKTFDNPQFADMSKETRGAPLISVP--LDWMLIYTRRYEANSILQNL 429  
DB 441 QNIVFHNKVPAGSN---ADWQRHFRDQRMILTPSDGLDRAVIAPORNHEHRLTLDL 497  
QY 430 FKVTTPANGMQRK-ALMEVDDRTREALVRLQOKVTAADTQIVCLSSNNKDKYDAIKKY 488  
DB 498 YRASGNGMLRIRSDQEFITIDRTGYVRAMDDCVSRDPKILILCLVNDNAERYSSIKR 557  
QY 489 LCTCCFPSPCCVAVRTLGKQOTVAIAIKALQNNCKSGELMEVDIPLKLVIMVIGDXY 548  
DB 558 GYVRAVFTQVYVLTKTCKPYSLMSIAIKAIQNNCKLGYPMIEIPLSGIMTIGDIA 617  
QY 549 HDMTAGRSIAGFVASIN-EGMTKRFSCITFODGQELVQGLKVCQLQALIRANNSCNEYM 607  
DB 618 KSTRDKRAVYALIASMDLQONSTYFSTVTECSAFDVANLTMIAKALRQYHERKYL 677  
QY 608 PSRIIVYRDVGDGQGLKTLVNVYVPOFLDCKSIGRGY-----NPLRVIVYKRVNTR 661  
DB 678 PSRIIVYRDVGDGQGLKTLVNVYVPOFLDCKSIGRGY-----NPLRVIVYKRVNTR 734  
QY 662 FFAOSGGLQNPFLPQTVIDEVETRPEDVDFIVSGAASGSVPYHNVIVIDNSGLKPDH 721  
DB 735 FFLNG---QNPPEGTIVDVITLPERYDFYVLSQVQGVTSPTSYNVLISNGLSDEK 790  
QY 722 IORLVYKCHITYYNMPGVIRVAPQVYAHKLAFLVQGSIHREBNLSINRLYYL 775  
DB 791 NQKLVYKCHITYYNMPGVIRVAPQVYAHKLAFLVQGSIHREBNLSINRLYYL 844

RESULT 12

AB862084  
ID AB862084 standard; protein; 843 AA.

XX AC AB862084;

XX DT 26-MAR-2002 (first entry)

XX XX Drosophila melanogaster polypeptide SEQ ID NO 13044.

XX XX Drosophila; developmental biology; cell signalling; insecticide;

XX XX pharmaceutical.

XX XX Drosophila melanogaster.

XX XX WO200171042-A2.

XX XX 27-SEP-2001.

XX XX 23-MAR-2001; 2001WO-US009231.

XX XX 23-MAR-2000; 2000US-0191637P.

XX XX 11-JUL-2000; 2000US-00614150.

XX XX (PEKE) PE CORP NY.

XX XX Venter JC, Adams M, Li FMD, Myers EW,

XX XX WPI; 2001-656860/75.

XX XX N-PSDB; ABL06187.

XX XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX XX genes from Drosophila and for elucidating cell signaling and cell-cell

XX XX interactions.

XX XX Disclosure; SEQ ID NO 13044; 21pp + Sequence listing; English.

XX XX The invention relates to an isolated nucleic acid detection reagent

XX XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX XX useful in developmental biology and in elucidating cell signaling and

XX XX cell-cell interactions in higher eukaryotes for the development of

XX XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX XX discloses genomic DNA sequences (AB1616176-AB1616177), expressed DNA

XX XX sequences (AB101840-AB101841) and the encoded proteins (AB357737-

XX XX AB357737). The sequence data for this patent did not form part of the

XX XX printed specification, but was obtained in electronic format directly

XX XX from WIPO at ftp://pub.int/pdb/published\_pdb\_sequences

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX Sequence 843 AA;

XX Query Match 34.8%; Score 1428.5; DB 4; Length 843;

XX Best Local Similarity 38.0%; Pred. No. 1.4e-129;

XX Matches 294; Conservative 163; Mismatches 284; Indels 33; Gaps 14;

QY 17 SKTSSGIIYRLSTNHFRLTSRPMALYOYHIDYPLMEARLRSLFQHEDLIGKCHA 76  
DB 88 SKKGTGVPMVLMQTNFRLKTPKEMRIYHVFESPIENPRKGVLSNANLIGSGL 147  
QY 77 FDGILFLPKLQOKVTEVSKTRNGEDVRIITITLNLPTSPCLQFYNIIFRLIKI 136  
DB 148 FDGLQFLTRKFEQETIVLSGSKLDIEYKISIKFVGFSICAPRFLQVNLILRSKMG 207  
QY 137 KNLQOIGRNYNPNDDIPSRHLVYPGFTSIILOYENSIMLCITVSHKVLRSFVDF 196  
DB 208 LNLVLGRNLPDPRAKIEIRPEKMEIMPGETSIRKHEKIDILGTETIKVMTETIYI 267  
QY 197 MENFYQTEBEHKQEOVSKELIGVYLTKNKNTYRVDDIDWQNKSTFKKADGSEVSF 256  
DB 268 MRCCHNPARH--QDEVKRVNVDLIVITDNNRTYRINDVDFQGTFRKSTF--SCKGRDISF 324  
QY 257 LEYRKQYNOEITDLKQPVIVSOPKRRRGGTLPKPMILPILCYITGILDMKRDENV 316  
DB 325 VEYLTKNYRIRIRHNPPLISK-NRDKALKTNASELVAVIPELCRVGTGLANMERNFOL 383

```

QY 317 MCDLAHTRLTPROREVGRLIDYTHKNDNV---RELADMGSLSPDNLSPSGRILOT 373
D 384 MRAMSSYTRNMPKOR---TDRLAFNRHLQNTPESSVVLDDNMNMLDKNTEVQGRILIGQ 440
QY 374 EKI--HOGGKTFDYNPQFADMSKETRGAPLISVYP--LDNMLLYTRRNYEANSILQNL 429
D 441 QNTVFHNGKVPAGEN---ADMQRHFRDQRMULTPSDGLDRWAVIAPQRNSHELRITLDSL 497
QY 430 FKVTAMGMQMRK-ALMIEVDRTETAYLRVLOQKVTADTQIVLCLLSNRKDKYDAIKKY 488
D 498 YRAASGMGLRIRSFQEFIIYDDRGTGYVRAMDVCVRSBPKLICTLVENDNAERISIKR 557
QY 489 LCTDCTPSPCCVAVARTGKQOTVMAITKIALQNCNCKMGSELNRVDIPLKLVNIVGIDCY 548
D 558 GYVDRAVPTQVYTLKTT-KRSLMSIATKIALQNLCLGYTPMKMIELPLSGMTIGFDIA 616
QY 549 HDMTAGRSIAGFVASIN-EGMTWFSRCIFQDRGQELVDGLKVCLOALRANNSCNEYM 607
D 617 KSTRDKRAVAGALIASMDLQNSTYFSTVTECSAFDVLANTLMEMIKALROYOHEHRKL 676
QY 608 PSRIIVRDGVGDQGLKTLVNYEVPQFLDCLKSIGRGY-----NPLRTIIVYKRVNTR 661
D 677 PSRIIVRDGVSSGLQFLFEFVKDIEKLT---EYARVQSPPLATIVVTRSWNTR 733
QY 662 FFAOSGRLQNPDPGTVIDEVTREPEWYDFPIVSQAVRSGSVSPTHNVIVDSGLKPDH 721
D 734 FFLNG---QNPPTIIVDVITLPERYDFYLSQVQVGTSPSTSYVNLVSSMGLSPEK 789
QY 722 IORLTYKCHLYNMPGVIRVPACQYAKLAFLVQGSIHREPULSINRLYYL 775
D 790 MOKLTYKCHLYNMSGTTRVPACQYAKLAFLVGNLHSLIPQNALKKFYTL 843

RESULT 13
AA90233
ID AA90233 standard, protein; 843 AA.
XX
AC AA90233;
XX
DT 29-AUG-2000 (first entry)
XX
DE Drosophila piwi protein.
XX
KW piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dytropy;
XX
KW anaemia; immunodeficiency; male infertility; Drosophila.
XX
OS Drosophila sp.
XX
FH key location/Qualifiers
FT MISC-difference 13 /label= Leu, Ile
FT MISC-difference 106 /note= "encoded by NTT"
FT MISC-difference 106 /label= Leu, Ile
FT MISC-difference 785 /note= "encoded by NTA"
FT MISC-difference 785 /label= Leu, Ile
FT MISC-difference 785 /note= "encoded by NTC"
XX
XX WO200032039-A1.
XX
XX PD 08-JUN-2000.
XX
XX PE 03-DEC-1999; 99MO-US028764.
XX
XX PR 04-DEC-1998; 98US-0110901P.
XX
XX PA (UYDU-) UNIT DUKE.
XX
XX PI Lin H;
XX
XX WPI; 2000-412085/35.
XX
XX N-PSDB; AAA07586.
DR

```

```

XX
PT piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and diagnostic
PT applications.
XX
PS Claim 4, Page 171-175, 201pp, English.
XX
CC This sequence represents the Drosophila piwi family protein, designated
CC piwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat tissue
CC dystrophy, anaemia, immunodeficiency, and male infertility
XX
SQ Sequence 843 AA;

```

```

Query Match 34.5%; Score 1416.5; DB 3; Length 843;
Best Local Similarity 37.7%; Pred. No. 26-128;
Matches 292; Conservative 163; Mismatches 286; Indels 33; Gaps 14;

```

```

QY 17 SKTSSGIIVRLSTNHFRLTSRPPQALYOHIDYPMARLRBALPQHEDLIGKCHA 76
D 88 SKKGTGVPMQLQTFEFAKTRPEWRIYHVEFPTLENPRVGVSNHANLSSGYL 147
QY 77 FPGTILFLPKRLQOKTEVFSKTRNGEDVRITITLNELPPTCLQFYNIIFRLIKI 136
D 148 FDLQLFTTRKEQELTVLSGSKDIEKISIKFVGFISCAEPRLQVLTNLRSMKG 207
QY 137 MNLQIGRNYNPNDBIDIPSHRLVIMPGFTTSILOYSIMLCTDVSHKVLRSVTLDF 196
D 208 LNELVGRNLPFRAKIRIFEFMELMPEYTSIRHEDDILGHEITRKVKTEYIYI 267
QY 197 MEFVYQTEHKEQEOVSKELGLVYLTKYNNKTYVDDIDMDQNKSTFFKADGSEVSF 256
D 268 MRCSNHPAH--QDEVAVNVDLIVLTIDYNNRTIYINDVDQGPYKSF--SGKGRDISF 324
QY 257 LEYVRKQNOETIDLKQPVLSQPKRRRPGGLTPGAMILPELCYLTGLTDRNDENV 316
D 325 VEYILTKNIRIRIDNHQPLISK-NRDKALKTVASELVILPELCVGTGIMEMERNFOL 383
QY 317 MCDLAHTRLTPROREVGRLIDYTHKNDNV---RELADMGSLSPDNLSPSGRILOT 373
D 384 MRAMSSYTRNMPKOR---TDRLAFNRHLQNTPESSVVLDDNMNMLDKNTEVQGRILIGQ 440
QY 374 EKI--HOGGKTFDYNPQFADMSKETRGAPLISVYP--LDNMLLYTRRNYEANSILQNL 429
D 441 QNTVFHNGKVPAGEN---ADMQRHFRDQRMULTPSDGLDRWAVIAPQRNSHELRITLDSL 497
QY 430 FKVTAMGMQMRK-ALMIEVDRTETAYLRVLOQKVTADTQIVLCLLSNRKDKYDAIKKY 488
D 498 YRAASGMGLRIRSFQEFIIYDDRGTGYVRAMDVCVRSBPKLICTLVENDNAERISIKR 557
QY 489 LCTDCTPSPCCVAVARTGKQOTVMAITKIALQNCNCKMGSELNRVDIPLKLVNIVGIDCY 548
D 558 GYVDRAVPTQVYTLKTT-KRSLMSIATKIALQNLCLGYTPMKMIELPLSGMTIGFDIA 616
QY 549 HDMTAGRSIAGFVASIN-EGMTWFSRCIFQDRGQELVDGLKVCLOALRANNSCNEYM 607
D 617 KSTRDKRAVAGALIASMDLQNSTYFSTVTECSAFDVLANTLMEMIKALROYOHEHRKL 676
QY 608 PSRIIVRDGVGDQGLKTLVNYEVPQFLDCLKSIGRGY-----NPLRTIIVYKRVNTR 661
D 677 PSRIIVRDGVSSGLQFLFEFVKDIEKLT---EYARVQSPPLATIVVTRSWNTR 733
QY 662 FFAOSGRLQNPDPGTVIDEVTREPEWYDFPIVSQAVRSGSVSPTHNVIVDSGLKPDH 721
D 734 FFLNG---QNPPTIIVDVITLPERYDFYLSQVQVGTSPSTSYVNLVSSMGLSPEK 789
QY 722 IORLTYKCHLYNMPGVIRVPACQYAKLAFLVQGSIHREPULSINRLYYL 775
D 790 MOKLTYKCHLYNMSGTTRVPACQYAKLAFLVGNLHSLIPQNALKKFYTL 843

```

RESULT 14







Wed May 12 09:50:22 2004

us-10-043-774b-2.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:06:35 ; Search time 17 Seconds  
(without alignments)

2373.786 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVNTRQNDHVKESKGT.....VQSIRHSPNLSLNRYLL 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1428.5	34.8	843	1	PIWI_DROME
2	557.5	13.6	834	1	IF2C_SGPO
3	510	12.4	1048	1	AGOI_ARATH
4	508.5	12.4	860	1	I2C3_MOUSE
5	506.5	12.3	857	1	I2C1_HUMAN
6	504.5	12.3	860	1	I2C2_RAT
7	504.5	12.3	861	1	I2C2_MOUSE
8	503.5	12.3	840	1	I2C2_RABIT
9	503	12.3	988	1	PIWI_ARATH
10	502	12.2	857	1	I2C3_MOUSE
11	499.5	12.2	860	1	I2C1_MOUSE
12	497	12.1	860	1	I2C3_HUMAN
13	495.5	12.1	861	1	I2C3_HUMAN
14	495.5	12.1	997	1	AGOI_ARATH
15	487	11.9	851	1	I2C2_HUMAN
16	421	10.3	1040	1	Y043_CAEEL
17	280.5	6.8	945	1	GCC7_CAEEL
18	248.5	6.1	1032	1	Y053_CAEEL
19	145	3.5	773	1	YD21_MRTUA
20	137.5	3.4	1912	1	DICE_HUMAN
21	134.5	3.3	1906	1	DICE_MOUSE
22	118	2.9	1120	1	MED_RICPR
23	114	2.8	2035	1	HMP2_YEREN
24	113.5	2.8	560	1	PHAC_PSEOL
25	113	2.8	949	1	PMAG_ARATH
26	113	2.8	3144	1	VP13_YEAST
27	111.5	2.7	1909	1	DICE_ARATH
28	108.5	2.6	1369	1	TRAP_HUMAN
29	108	2.6	1317	1	N145_YEAST
30	108	2.6	1653	1	CIH_YEAST
31	107	2.6	1282	1	SN3X_MOUSE
32	106.5	2.6	690	1	Y173_UNEPA
33	104.5	2.5	1966	1	DOC4_HUMAN

34	104	2.5	1180	1	C4AA_BACTI	P16480 bacillus th
35	104	2.5	1277	1	PDS5_YEAST	O04264 saccharomyc
36	104	2.5	4128	1	PRKD_MOUSE	P97313 mus musculu
37	103	2.5	1342	1	RPOB_HAERIN	P43738 haemophilus
38	103	2.5	1575	1	I062_HUMAN	O13576 homo sapien
39	102.5	2.5	1343	1	CHLI_CHLVU	P26304 chlorella v
40	102.5	2.5	869	1	YD95_YEAST	P38966 saccharomyc
41	102.5	2.5	1170	1	IT2A2_BOVIN	P53710 bos taurus
42	102	2.5	3122	1	DEOZ_MOUSE	O61493 mus musculu
43	101.5	2.5	3135	1	S230_PLAFO	O08372 plasmodium
44	101	2.5	1388	1	RPOD_TOBAC	P38550 nicotiana t
45	101	2.5	1886	1	POL_COYIV	P19109 commelina y

#### ALIGNMENTS

RESULT 1  
PIWI\_DROME STANDARD; PRT; 843 AA.  
ID PIWI\_DROME  
AC O9VXMI; 096674; 096675;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE PIWI protein  
GN PIWI OR CG6122.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
[2]  
SEQUENCE FROM N.A., FUNCTION, DEVELOPMENTAL STAGE, AND TISSUE  
SPECIFICITY.  
RC STRAIN=Canton-S, and Oregon-R.  
RX MEDLINE=99069219; PubMed=9851978;  
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.,  
RT "A novel class of evolutionarily conserved genes defined by piwi are  
RT essential for stem cell self-renewal."  
RL Genes Dev. 12:3715-3727(1998).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA Amanatides P.G., Scherer S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chang Q., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.M.,  
RA Beeson K.Y., Bemis P.V., Berman B.P., Bandaru D., Bolshakov S.,  
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostel R., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey K.A., Howard T.U., Hernandez J.R., Houck J.,  
RA Hootin D., Houston K.A., Howard T.U., Wei M.-H., Idegawa C.,  
RA Jatalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laaso P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

QY	25	LEYYKRONOELTDLKQFVAVSQPKRRRGGLPGPAMILPELCYNGYVDMKRDENV	313
Db	325	VEYLLTKNINIRHNOGPLLISK-NRDKAKTVNASELVVILPELCVGTGIMNEMRNFOI	383
QY	317	MDCLAVHTRLPPEQOREVGLIYIKHNDNVQ---RELADMLSPDSNLSLSPSGIILOF	373
Db	384	NKAMSSYTRMPEKOR---TDRLRAFNRLONTPEBSVAVLSDMMNEIDKQVTEVOGHIIOQ	440
QY	374	EKI--HOGKTPDYVPPFADSSKETRGPALISVPE--LDNMLLIYTRRNEAANSIIQNL	429
Db	441	QNTIVHNGKVPGEN---ADWQRFHPRDGMMLTTPSDGLDWMVAIYQPRNSHELRLLDLSI	497
QY	430	FXYTPAMGOMKMK-ALIMEVDDRTREAIRVLQCKYTDADTQIVCLLSNRKDKYXDIKKY	488
Db	498	YPRASGMGRIRISPOEFIIYDRKGTGYVRANDCVRSDDPLCLLPNDNAEYSSIKKR	557
QY	489	LCNDCTPSPQCVVATLKCOOTVAAIATKIALQNNCKSGMGBELRWVDIPLKLVITGIDCY	548
Db	558	GYDRAVPEQVYVTLKTI-KNSMSIAKXIALQNCCKLGTPWMIELPLSGMLTIGDIA	616
QY	549	HDMTAGRSIAGFVASIN-EGMTREFSCITQDNGQGLNGLKVCQALAFANNSCNEYM	607
Db	617	KSTRDRKRYGALIASMDLQONSTYFSTVTECSAFDYALNTIMIKAKLRQVQHEHRL	676
QY	608	PSRIIYRVDGVDGDLKTLWNEVEYQFLDCKSIGRGY-----NPLTIYVKKGVNTR	661
Db	677	PSIIVRYRQVSGSLKQLEFEVMDIIIEKLT--EYARVQLSPPOLAVIYVTRSMNTR	733
QY	662	FFKQSGRLQNLPLPGVIVDEVTRERWDEFTVQAVRSGSVSTHNVIYDNGLKEPDH	721
Db	734	FFLNG---QNPFGTIVDVTLTERDPEFVASQVACGVSTSTINVLSSMGLSPEK	789
QY	722	IGRLTYKLCHIYVNPVGIVRVAPCOYAKLIAFLVQGSIHREPNLSLNRLYYL	775
Db	790	MOGLTITMGLLYNMNGTTRVAVQVAKKLATLVGTMHISIPONALEKKYYL	843
RESULT 2			
IF2C_SCHPO	ID	IF2C_SCHPO	STANDARD; PRT; 834 AA.
AC	074957/		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DS	Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).		
GN	SPCC736.11.		
OS	Schizosaccharomyces pombe (fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OK	NCBI_TaxID=4896;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RX	MEDLINE=21846401; PubMed=11859360;		
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	Agours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,		
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,		
RA	James K., Jones L., Jones M., Leather S., McDonald S., Molten J.,		
RA	Mooney P., Moule S., Mungall K., Murphy L., Nibbelte D., Odell C.,		
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,		
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	Skellern J., Simmonds M., Squares R., Squares S., Stevens K.,		
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	Woodward J., Volckaert G., Aert R., Robben J., Gilmontrez B.,		
RA	Welford J., Vanstreels E., Rieger M., Schaefer M., Meller-Thier S.,		
RA	Gabel C., Fuchs M., Fritzc C., Holzner E., Moesl D., Hilbert H.,		
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,		
RA	Goffeau A., Cadieu E., Drenth S., Gloux S., Lelure V., Mottier S.,		

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 Lucas M., Rocher M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cernetti L., Lowe T., McComble W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nure P.,  
 RT "The genome sequence of *Schistosoma* *pompe*."  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain  
 CC initiation process (By similarity).  
 CC -1- SIMILARITY: Belongs to the argonaute family.  
 CC -1- SIMILARITY: Contains 1 PAZ domain.  
 CC -1- SIMILARITY: Contains 1 Pwi domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.embnet.org/licenses/embnet.html> or send an email to [license@embnet.org](mailto:license@embnet.org)).  
 CC -----  
 CC EMBL/AL023705; CAA19275.1; -  
 DR PIR; T41568; T41568.  
 DR GenBank; SP000571; C:chromatin; ISS.  
 DR GO; GO:0005717; C:chromatin; ISS.  
 DR GO; GO:0005694; C:chromosome; ISS.  
 DR GO; GO:0016283; C:eukaryotic 48S initiation complex; ISS.  
 DR GO; GO:0005654; C:nucleosome; ISS.  
 DR GO; GO:0005718; C:nucleosome; ISS.  
 DR GO; GO:0007049; P:cell cycle; ISS.  
 DR GO; GO:0016943; P:cell organization and biogenesis; ISS.  
 DR GO; GO:0006335; P:establishment and/or maintenance of chromatin; ISS.  
 DR GO; GO:0006412; P:protein biosynthesis; ISS.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.  
 DR GO; GO:0006350; P:transcription; ISS.  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; Pwi.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; Pwi; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; Pwi; 1.  
 KW Initiation factor, Protein biosynthesis.  
 FT DOMAIN 212 326 Pwi.  
 FT DOMAIN 500 799 Pwi.  
 SO SEQUENCE 834 AA; 94438 MW; FSCB397BF590D95A CRC64;  
 Query Match 13.6%; Score 557.5; DB 1; Length 834;  
 Best Local Similarity 22.9%; Pred. No. 2,4e-32;  
 Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;  
 QY 20 GSSGIVLSTNNPFLTRPQWALYOHYIDNPMLERARLSALLFQHEU-----70  
 DB 16 GGLGQITLKANFPQILSPNETTINQYIVGDSRVPRKSSQLIMSKVYKQYFGSSSM 75  
 QY 71 -----IGKCAFPDGTILF-----LPKRLQ--QKTEV-----FSTRNGED 104  
 DB 76 NSVYDGRSMCKSGDIADGTIKVNGSHPREIFSSIOKSSKINLHTLSQFVNSKXSSD 135  
 QY 105 VRIITTLNEPSPPTGLQFNNIFRR-----LLKINNLQIQIGNRYNPNPDIDISHR 159  
 DB 136 -----PQVLSSIMFDLLKXKXKXSETLFGFMHSPFTGNGVSLGCGV-----176  
 QY 160 LVIMPFGTTSILQYENSIMLCTDVSHKYL-RSEYVLDPM-----197  
 DB 179 --AMGFGVQISIPNGFMSVNVNDISSAFMRNDISLQIMLEYTDCSNVRDLTRFDLRLS 236  
 QY 198 --FNHYHQTEHKKFPEQVSKELIGVYLTKNNKTYRVDLIDWQNPSTF-KKADGSE- 253  
 DB 237 RFRRLKVTYQHRNN-----VGIDLANKRYSTIEGSSKASASFPYRLNGSEQ 285  
 QY 254 -VSFLYRKYQYNGEITDLKQVPLVSPKRRRPGGTLPGPAMLIPELCYLT--GLTDK 309

DB 286 KISVAEYLENNVRLQYFNPILPCL-----VNGAMLPDIEFCVYKQRYTAK 333  
 QY 310 MENDFNWKDLAVHTRLPPEQOREVGLIDYHKND-NVORELMDGLSPSNLSBSG 368  
 DB 334 LNSD-----QPMNIRFVAVQSPFERVQOIDDVHMDWDTDPYLYQYQKTKQKMLEVPA 388  
 QY 369 RLOTEKHIGGKPTDVPNPQPADNSKETRGAPLISV-KPLDNLILI--YTR-NEYAA 422  
 DB 389 RYLETPSIRYGGDCLF-RVSGRMV-IRGRFLDPRAPIRSMVMTSTIRLRPMGI 445  
 QY 423 NSLIONLFRVTPAMQO--MRKAMIEVDNR--TEAYRVLQKY--TADTQIVCLL 474  
 DB 446 ENFLQTVQTLTSLGIDNFMKKPPLVADIRSGVELCTILYKKAEOVGANAPPDYLFFIL 505  
 QY 475 SSNRDKDAIKYLCCTDPTSPGCVARTLGKQOTWALATKIALONCKMG--ELM 531  
 DB 506 DKSSEPPGSIKRVGNTMLGVPSQAIKSHI--LQSKFOYCANLGMKINVKVGINCLLI 563  
 QY 532 RVDIPKLY--MIVGIDCYH-DMTAGRSIAGFVASINEGMTRMPSRCIFDQGGELVDG 588  
 DB 564 PKNPLGANPITLILGDVYHPGVANGVSIASIVASVLDNGCKYAVGSRQPRHQEVLIG 623  
 QY 589 LVYCLQALIRANVSCNEMVPSRIIVYRGVGDQLKTLVNVVPPFLDCLSGISGNPR 648  
 DB 624 MDIVVYLLQGFRAMTKQPPRIIVFRDGTSGQLSVINDELSQLKACHSLSPKPNK 683  
 QY 649 LVYIVVKKRVNTRPFAQS--CGRLONPLPGTVIDVEYTRPEMYDFFIYVQAVSGSVSP 705  
 DB 684 ILVCTQKSHHRRFFIKNSDSDRGNPLPGTILIKRYTHPQYQYFYLISHSLGVSVP 743  
 QY 706 THYNYLYDNGSKPDIHQILYKLCIIYNNMGVIRVPAPOYAKKAFI-----755  
 DB 744 VHYTVLHDEIRQMPDQFQTLQYVYATSAVSLVPYVYHVLVSLARYQDVTAD 803  
 QY 756 -----VGQSIHREPLSISNRL 772  
 DB 804 TVETSEASMDQEV-KPLALUSSKL 827  
 RESULT 3  
 AGOI ARATH STANDARD; PRT; 1048 AA.  
 ID AGOI ARATH STANDARD; PRT; 1048 AA.  
 AC 004379; Q9UP83;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Argonaute protein.  
 GN AGOI OR AT1G48410 OR P1A17.3 OR T1N15.2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eucoisids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RC MEDLINE=21016719; PubMed=1130712;  
 RA Theologis A., Becker U.R., Palm C.U., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Malt R., Marshall A.,  
 RA Malt R., Malt R., Nguyen M., Niem W.C., Osborne B.I.,  
 RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salberg S.J., Schwartz J.R., Shin P., Southwick A.M.,  
 RA Sun H., Tallon L.V., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Uteback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RU Nature 408:816-820(2000).  
 RA [3]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Timp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Atrakawa T., Bano J., Banno F., Bowser L., Brooks S.Y., Cathinici P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Saou M., Tame R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,  
 RT "Empirical analysis of transcriptional activity in the Arabidopsis  
 genome.";  
 RL Science 302:842-846(2003).  
 CC -1- FUNCTION: Essential for proper development of leaves and floral  
 organs and formation of axillary meristems.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Even=Alternative splicing; Named isoforms=1;  
 CC Comment=Alternative number of isoforms are produced. According to EST  
 sequences;  
 CC Name=1;  
 CC IsoId=004379-1; Sequence=displayed;  
 CC Tissue specificity: Widely expressed at low levels.  
 CC -1- DEVELOPMENTAL STAGE: Expressed throughout all developmental  
 stages.  
 CC -1- SIMILARITY: Belongs to the argonaute family.  
 CC -1- SIMILARITY: Contains 1 PAZ domain.  
 CC -1- SIMILARITY: Contains 1 PIWI domain.  
 CC -1- CAUTION: Ref.2 (AA079718) sequence differs from that shown due to  
 erroneous gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC DR EMBL, U91995; AAC18440.1; -;  
 CC DR EMBL, AC007932; AAD9755.1; -;  
 CC DR EMBL, AC020889; AAF79718.1; ALT\_SEQ.  
 CC DR EMBL, BT000941; AAN1341.1; -;  
 CC DR InterPro: IPR003100; PAZ.  
 CC DR InterPro: IPR003165; PIWI.  
 CC DR Pfam: PF02170; PAZ; 1.  
 CC DR Pfam: PF02171; PIWI; 1.  
 CC DR PROSITE, PS50821; PAZ; 1.  
 CC DR PROSITE, PS50822; PIWI; 1.  
 CC DR Developmental protein; Alternative splicing.  
 CC FT DOMAIN 391 501 PAZ.  
 CC FT DOMAIN 676 997 PIWI.  
 CC FT DOMAIN 13 104 GLY-RICH.  
 CC SEQUENCE 1048 AA; 116190 MW; 3E514633A09C41 CRC64;  
 Query Match 12.4%; Score 510; DB 1; Length 1048;

Best Local Similarity 25.1%; Pred. No. 9,2e-29;  
 Matches 213; Conservative 136; Mismatches 376; Indels 124; Gaps 33;  
 QY 20 GSSGIIVRLSTNHRRLSPQWALYYOHIDYNPMEARLRSLPQHEDLIGCH---- 75  
 DB 186 GSGKRCIVANHF-FAELPDKLHHYDVTTEVTSRGVAVMKQLVNYSRSHGSR 244  
 QY 76 -ARDG-----TIFLP-----KLOOKTEVSKTRNGEDVRII 109  
 DB 245 LPAYDGRKSLYTAGLPFPNSKEPFINLLDEVGAGGRRRRFRVAVIKVARDLHILGM 304  
 QY 110 TLTELPPTSPCTCFYNIIFRRLKIMNLOQIGRNYNPNDPIDPSH-----LVWP 164  
 DB 305 FLEKQSDAPQEAQVADIVLRE-LPTRYIPVGRSFYSP-----DIGKQSLDGLSWR 359  
 QY 165 GFTTSILOYSINLCTDVSHKV-LRESEVLDFWENFYHOTBEHK-----FOEVSKELI 218  
 DB 360 GFYOSIRPTQGLSINIDMSSTAFIEANPVYQFCDLLNRDISRPLSDADRVYIKALR 419  
 QY 219 GL-VLTKNK--KTYRVDDIDMDQNPKEFKAD-GSEVSFLERYKQYNOEITDJKOP 274  
 DB 420 GYKVEYTRGMRKRYISGLTAVATRELTPVDERNTQSVVEYFHEYGFRIQHTOLP 479  
 QY 275 VLVSQPKRRRPGGTLPEPAMLPELCYLT--GLTDKRDENVMKDLAVHTRILPBRQ 332  
 DB 480 CLQVGNNSR-----PNYLPMEVCKIVGQRYSKRLRQITALLKV-TCQRPIDRE 529  
 QY 333 REV---GRLLIDYHNDVQSELDWGLSPSNLSFSGRIQTE--KIHQSGTFDYNP 387  
 DB 530 KQILQTVOLNDY--AKDYVADE--FGIKISTLSAYEARLTPPWLKYHSGREGTCLP 584  
 QY 388 QFADWSKETRGAPLISYKPLDNWLLIYTRYNE--AANSLIQNFKYTPAMGMQR-KAI 444  
 DB 585 QVGQWNNMNR--KMINGTVANNWICINFSRGVODNLATFOQLAQWQYSGKAFNPEPV 642  
 QY 445 MIEVDRTREVALRYLQCKVTADT-----QIVCLLSNRRKDYAIKKYLCCTDPT 495  
 DB 643 LPVSAREQVEKVKTKYTHDATSLSQGEKIDLLIVLPNNNSLVGDDLRIETELGI 702  
 QY 496 PSQCVAARTLQK-QGVNAIAIKALQWCKMG-ELMRV-----PL--KVMYIGI 545  
 DB 703 VSQCLTGVFPMKSYQA--NVALKINIVYGGHNTLVVALSRILPVSDDPTIRGA 759  
 QY 546 DCTHDMTA--GRSLIAGVYASIN-EGMTRWFSRCIFDORGGELVDGL----- 589  
 DB 760 DVTTHHPGSDSPSIAAVASQDPETIKYGLVCAQHRQELIDDLKEMKDPKGYVT 819  
 QY 820 GGMTELLIAFPKRSQGHKPLRIIFPDVSGEQFQVLYVELDIAIRKACASLEADYQPPV 879  
 QY 650 YIVYVKKRNTFFAO-----SGRLNPPLPGYIVDEVRPEVYDFIVSQAVRSGSV 703  
 DB 880 TVVVYQKRHTLRFQNHNDHSVDRSGNIPGLYVDSKICHPTEPDYLSCHADIGQS 939  
 QY 704 SPTHVIVYDNGSLKPDHIOQLTYKLCIHYNNPQVIRVAPDCQYAHKLAFLVGSIHRE 763  
 DB 940 RPAHVHVLVDENNFTADGQSLTNMLCYTACRSVSVIPPAYVAHAAFPRA--RFYWE 997  
 QY 764 PNLISNRL 772  
 DB 998 PETSDESM 1006  
 RESULT 4  
 ID12C2\_MOUSE STANDARD; PRT; 860 AA.  
 AC Q8CUG0;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)  
 DE (PWI/argonaute family protein mef2C2).

EN E1P2C2. . . . .

OS Mus musculus (Mouse) .

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RM [1]

RP SEQUENCE FROM N.A.

RA Doi N., Zanno S., Yi-Tel K., Takahasi F., Ueda R., Miyata Y.,

RA Saigo K.,

RT "elf2C family proteins and Dicer homologs are required for siRNA

RT mediated RNAi in mammalian cells."

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain

CC initiation process (By similarity).

CC -1- SIMILARITY: Belongs to the argonaute family.

CC -1- SIMILARITY: Contains 1 PAZ domain.

CC -1- SIMILARITY: Contains 1 P1w1 domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC

DR EMBL; AB081472; BAC15767.1; .

DR WGD; MG1;2446632; E1P2C2.

DR InterPro; IPR003100; PAZ.

DR InterPro; IPR003165; P1w1.

DR Pfam; PF02170; PAZ; 1.

DR Pfam; PF02171; P1w1; 1.

DR PROSITE; PS50821; PAZ; 1.

DR PROSITE; PS50822; P1w1; 1.

DR Initiation factor; Protein biosynthesis.

FT KMT 236 349

FT DOMAIN 236 349 PAZ.

FT PT 518 819

SQ SEQUENCE 860 AA; 97294 MW; DA32A759EA467B31 CRC64;

Query March 12.4%; Score 508.5; DB 1; Length 860;

Beet Local Similarity 23.8%; Pred. No. 8.8e-29;

Matches 195; Conservative 152; Mismatches 350; Indels 121; Gaps 32;

QY 20 GSSGIIVLSTNTRLTSLRPMALYQYHIDNPLMEARLRSL--LFOH--EDLIG-K 73

DB :::::V:::L:::T:::R:::L:::S:::P:::M:::A:::L:::Y:::Q:::Y:::H:::I:::D:::N:::P:::L:::M:::E:::A:::R:::L:::R:::S:::L:::--:::L:::F:::O:::H:::--:::E:::D:::L:::I:::G:::-:::K::: 73

DB 33 GTGRTIKLQGNPFEM-DIPKIDIHVELDIKPEKRPRVRNREIYEHVQGFNKQIFEDR 91

QY 74 CHADG-----TILFLPKLQOKTEVSKTRNGD--VRITL-----TLTN 113

DB :::::T:::I:::L:::F:::L:::P:::K:::L:::Q:::O:::K:::T:::E:::V:::S:::K:::T:::R:::N:::G:::D:::--:::V:::R:::I:::T:::L:::--:::--:::T:::L:::T:::N::: 113

DB 92 KPVDGRRKRLYTAAPLPTG-RDKYLEVTLRGEGDRILKSIKVCSCVSLQALHDAISG 150

QY 114 ELPEPTSPCLQFVNIIFRLILKIMLQOIGHYVNPNDPIP-SRLVIMPGFTTSILQ 172

DB :::::E:::L:::P:::E:::P:::T:::S:::P:::C:::L:::Q:::F:::V:::N:::I:::F:::R:::L:::I:::L:::K:::I:::M:::L:::Q:::O:::I:::G:::H:::Y:::N:::P:::N:::D:::P:::I:::P:::-:::S:::R:::L:::V:::I:::M:::P:::G:::F:::T:::S:::I:::L:::Q::: 172

DB 151 RLPSVPPEFTIGALDVNWH-LPSMRYPFVGRSFFLASGSCGNPLGGGSEVFGTHQSVR 209

QY 173 YENSIMLCTDVSHKYL-RSEYVLD--MNFYHQTENK----FQGVSKELIGVYL 223

DB :::::Y:::E:::N:::S:::I:::M:::L:::C:::T:::D:::V:::S:::H:::K:::Y:::L:::-:::R:::S:::E:::Y:::V:::L:::D:::--:::M:::N:::F:::Y:::H:::Q:::T:::E:::N:::K:::--:::--:::F:::Q:::G:::V:::S:::K:::E:::L:::I:::G:::V:::Y:::L::: 223

DB 210 SLMMQMLNIDVSAFYAQAQVIEVCEVLDFKSIIEQOKPLTDSORVYKFTKEIKGLAVE 269

QY 224 TKY---NNKTYRVDDIDWDQNPSTFKKADS---EVSFEYRYKQVNGEITLQKQVYL 276

DB :::::T:::K:::Y:::--:::--:::N:::N:::K:::T:::Y:::R:::V:::D:::D:::I:::D:::W:::D:::Q:::N:::P:::S:::T:::F:::K:::K:::A:::D:::S:::--:::--:::E:::V:::S:::F:::E:::Y:::R:::Y:::K:::Q:::V:::N:::G:::E:::I:::T:::L:::Q:::K:::Q:::V:::Y:::L::: 276

DB 270 ITHGGQMKRKRVCVNTRPASHQTFPLQOSSGQVCEVTAQVQYKDRHKLVLKRPHPCL 329

QY 277 -VSGPKRRRGEGTLPGPMLIPELCYLTG-----LTDKRRDNFVYMKDLAVHTLTP 328

DB :::::-:::V:::S:::G:::P:::K:::R:::R:::G:::E:::G:::T:::L:::P:::G:::P:::M:::L:::I:::P:::E:::L:::C:::Y:::L:::T:::G:::--:::--:::--:::L:::T:::D:::K:::R:::R:::D:::N:::F:::V:::Y:::M:::K:::D:::L:::A:::V:::H:::T:::L:::T:::P::: 328

DB 330 QVGEQOKA-----TYLPLEVCNIVAGQRCIKTLTDQGTSM-----IRATARSAP 374

QY 329 EGQREVERGLDIYHKNDNVQRELBDWGLSPDNLSTSGRLITQTEKHGGKGFDPVNPQ 388

DB :::::E:::G:::Q:::R:::E:::V:::E:::R:::G:::L:::D:::I:::Y:::H:::K:::N:::D:::N:::V:::Q:::R:::E:::L:::B:::D:::W:::G:::L:::S:::P:::D:::N:::L:::S:::T:::S:::G:::R:::L:::I:::T:::Q:::T:::E:::K:::H:::G:::G:::K:::F:::D:::P:::V:::N:::P:::Q::: 388

DB 375 D-ROBEISKLR--SASFTNDPIYERFQIMYDENTDVTGVLQPSLIVGR-----424

QY 389 FADMSKREKGAFLISVYPLDN-----WLLIYTRNRYEA--NSLIQNTFKYT 433

DB :::::F:::A:::D:::M:::S:::K:::R:::E:::K:::G:::A:::F:::L:::I:::S:::V:::Y:::P:::L:::D:::N:::--:::--:::--:::W:::L:::L:::I:::Y:::T:::R:::N:::Y:::E:::A:::--:::N:::S:::L:::I:::Q:::N:::T:::F:::K:::Y:::T::: 433

```

Db 425 -----NKAIAITPVGVWMDMRNKKOFTGIEIKWMAIACAPROCTEYHAKSFTBOLAKIS 479
Qy 434 PAMOMQRK-----AIMEVDVDRTEAYLRVLOQKQTADTQIVLCILSSNRKDKYDAIKYL 489
Db 480 RDAMPPIIOGQPCFEKVAQGADSVPEMRHL-KNTYAQLGLTVVIL-PGKTPVVAEVRKVG 537
Qy 490 CTDCPPTSSQCVANTLGGKQGTWMAIAFKIALQNNCKRGG-----ELMRDPIPLKVMIV 543
Db 538 DTVLGMATOCQOMKNV--ORFTPTQLNLCLIKINVKIGGVNNILLPGGRPEVFQGPVIFL 595
Qy 544 GIDCYHMTAG---RSISIAGVASINENGMTWFSRCIFPDROGELVGLKVCLOALRAW 600
Db 596 GADVTHP-PADGDKRPSIIAIVGSMDMAPRKVCATYVAQOHRQEIIDDLAAMRELLIQF 654
Qy 601 NSCNEMPSRIIVYRDGVGQGLKTLVNVYEPQFLDCKSIGRGYNRLTVIVYKRVNT 660
Db 655 YKSTRPFPRTIIFRDSDVSEGOFGQVLAHHEILAIRACICILKEXYQGIETFIIVQKRMHT 714
Qy 661 RFFA-----QSGGRQLNPLPGVIDEVTRPEWDFPVSQAVNSGSVPTHYVINYNSG 716
Db 715 RLFTCDKNERVGSNGNPAGTIVDTKTIHPPEDFYICSHAGIGQIGTERPSHYHVLMDNR 774
Qy 717 LKPDHIQRLTYKLCHIIYNMPPVIRVAPQCYAHKLAIF 754
Db 775 FSSDELQILFYQLCHTYVRCRBSVSIAPAYVALVAF 812

RESULT 5
ID 12C1_HUMAN STANDARD; PRT; 857 AA.
AC 09ULI8;
AD 16-OCT-2001 (Ref. 40, Created)
DT 16-OCT-2001 (Ref. 40, Last sequence update)
DT 28-FEB-2003 (Ref. 41, Last annotation update)
DT DE Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
GN (Putative RNA-binding protein q99).
OS EIF2C1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005943; PubMed=10534406;
RA Koesters R., Adams V., Betts D., Moos R., Schmid M., Sternann A.,
RA Haasman S., Weitz S., Lichter P., Heltz P.U., von Knobel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression."
RL Genomics 61:210-218 (1999).
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 Pw1 domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```



DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; PwI; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; PwI; 1.  
 DR Initiation factor; Protein biosynthesis.  
 FT DOMAIN 226 346  
 FT DOMAIN 515 816  
 FT DOMAIN 857 AA; 97214 MW; 1DBB524AE7CBAF66 CRC64;  
 SQ SEQUENCE

Query Match 12.3%; Score 506.5; DB 1; Length 860;  
 Best Local Similarity 24.1%; Pred. No. 1.7e-28;  
 Matches 203; Conservative 160; Mismatches 349; Indels 133; Gaps 37;

20 GSSGIVRLSTNHLRSLSPQWALYQYHIDYPLMEARRLSAL--LFQH--EDLIG-K 73  
 30 GTVGKPIKLANYEV-DIPKIDYHYEVDPKPCRRVREVEYVQHFQKQIFGDR 88  
 74 CHAFDQ-----TIFLP---KELQKQTEVPSKTRNGD--VRIT-----T 110  
 89 KPYDGKKNIYVATLPNGRNVDFEYV---IPGSKDRIKFKSITKLAIVSRMLHEA 144  
 111 LTNELEPTSCLOFYNIIFRLKIMLQOIGHYVNPNDPIDP-SHRLVIMPGETTS 169  
 145 LVSGQIPVPLESVQALDVAMHLS-KRYTPVGRSFPSPGGYHPLGGREWVGFHQS 203  
 170 ILQYENSIMLCTDVSHKVL-RSEYVLDPMF-----NFYQTEHKEQEQV--SKELIG 220  
 204 VRPAMMMKMLNIDVSAFAFYKQVIEFMCEVLDIRNIDPQKLTDSQVRFRFKEIKGL 263  
 221 VILRYK--NKKYRVDIDMDQNPSTF---KKADGSEVSEFLEYRKQYNOEITDLKQ 273  
 264 KVEYTHGQGMGRKRVCAVTRPASHQTFPLQESQGVETVAQYKQKYNQ--LKY 320  
 274 PVL---VSQPKRRRREGTLPGPAMLIPELCYLTG-----LTKMRDENVMKDLAV 322  
 321 PHLPCLAGQGEQKH-----TYLPLEVCNVAQGRCTIKKLTLD--NQTSTM--IKA 365  
 323 HTLRLPGRQREVGSLIDYIHKNDNVQRELDWGLSFDNSLSSGILQTEKHQGGKT 382  
 366 TARSAPD-KQBEISLHK--NASTNLDPIYIEFGIKVMDTENTYGVLPAPILQYGRN 422  
 363 PDY-NPQPADWSKETRGAPLISVXPLDMLIYFRNYSANSLIQ---LFCVTPAMG 437  
 423 RAIAIPNGQW--DMRGQFYNGEIKYMAIACFAPQCKREEVLYKNFTDLRKISDAG 480  
 438 MQMRK---AIMEVDDBTEAVLYRLQOKTADQIVVCLSSNRKQYDAIKYKLTDC 493  
 461 MPIQQCFCKYKQADSVENFRLKNTYSG-LDLIIVL--PKGTPIYAEVKGVDITL 538  
 494 PTPSQCVARTLQKQOTVMAIATKIALQMCNKGELWRVDP-----LKVMTVIGID 546  
 539 GWATQCVQKVVK--TSPTQISNCLKINYLG--INNIIIVPQRSVAFQOPIVIFGLAD 595  
 547 CYHMDTAG---RSLAGVVASINEGMTWFSRCIFQDQSGVLDGKVCQALAAIRANSC 603  
 596 VTHP-PADGKKRPSITAVAGSDAHPSRYCATVVRQPSRQSTIEBLSYMWELLIQYKS 654  
 604 NEYMSRIIVYDGVGQDLKTLVNYEVPQFLDLKLSIGRGYNPLTVIVYKKEVNTFF 663  
 655 TRFKETRIIFRYDGVPEGLPQILHYELLARDACTIKLEKDYQDPITIVVQKHHRLFF 714  
 664 A---QSGRLONPLPGYIVDVEVTRPEMYDFPIYSQAVRSGSVSPFTYNTYVNSGLKP 719  
 715 CADKNERIGKSNIDAGITVDNITNTHPEFDFYLSHAGIQTSRPSHYVLMNDNRFTA 774  
 720 DHIGRLTYQLCHIIYNNWPGVIRVPAFCQYAHKLAF-----LVGOSIH 761  
 775 DELQILITVGLCHTYVRCRVSISIPAPAYARLVARFARYHLVDKHDGSGSHISGQNG 834  
 762 REP 764  
 835 RDP 837

RESULT 6  
 12C2 RAT  
 ID 12C2 RAT  
 AC 090281;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)  
 GN (Go1g1 ER protein 95 kDa) (GERP95).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Hepaticoma;  
 RX MEDLINE=99443791; PubMed=10512872;  
 RA Chakaluk D.E., Tabhaz N., Hendricks L.C., DiMattia G.E., Hansen D.,  
 RA Pilgrim D., Hodman T.C.,  
 RT "GERP95, a membrane-associated protein that belongs to a family of  
 RT proteins involved in stem cell differentiation.",  
 RL Mol. Biol. Cell 10:3357-3372(1999).  
 CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain  
 CC initiation process.  
 CC -1- SIMILARITY: belongs to the argonaute family.  
 CC -1- SIMILARITY: contains 1 PAZ domain.  
 CC -1- SIMILARITY: contains 1 PwI domain.  
 CC -1- SIMILARITY: contains 1 PwI domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC  
 CC EMBL; AF195534; AAF12800.1; ALT INIT.  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; PwI.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; PwI; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; PwI; 1.  
 KM Initiation factor; Protein biosynthesis.  
 FT DOMAIN 226 346  
 FT DOMAIN 518 819  
 FT DOMAIN 860 AA; 97318 MW; A5B0798C66481C9C CRC64;  
 SQ SEQUENCE

Query Match 12.3%; Score 504.5; DB 1; Length 860;  
 Best Local Similarity 24.1%; Pred. No. 1.7e-28;  
 Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;

20 GSSGIVRLSTNHLRSLSPQWALYQYHIDYPLMEARRLSAL--LFQH--EDLIG-K 73  
 33 GTTGRTIKQANFFEN-DIPKIDYHYELDIPKPCRRVREVEYVQHFQKQIFGDR 91  
 74 CHAFDQ-----TIFLPKELQKQTEVPSKTRNGD--VRIT-----T 110  
 89 KPYDGKKNIYVATLPNGRNVDFEYV---IPGSKDRIKFKSITKLAIVSRMLHEA 144  
 111 LTNELEPTSCLOFYNIIFRLKIMLQOIGHYVNPNDPIDP-SHRLVIMPGETTS 169  
 145 LVSGQIPVPLESVQALDVAMHLS-KRYTPVGRSFPSPGGYHPLGGREWVGFHQS 203  
 170 ILQYENSIMLCTDVSHKVL-RSEYVLDPMF-----NFYQTEHKEQEQV--SKELIG 220  
 204 VRPAMMMKMLNIDVSAFAFYKQVIEFMCEVLDIRNIDPQKLTDSQVRFRFKEIKGL 263  
 221 VILRYK--NKKYRVDIDMDQNPSTF---KKADGSEVSEFLEYRKQYNOEITDLKQ 273  
 264 KVEYTHGQGMGRKRVCAVTRPASHQTFPLQESQGVETVAQYKQKYNQ--LKY 320  
 274 PVL---VSQPKRRRREGTLPGPAMLIPELCYLTG-----LTKMRDENVMKDLAV 322  
 321 PHLPCLAGQGEQKH-----TYLPLEVCNVAQGRCTIKKLTLD--NQTSTM--IKA 365  
 323 HTLRLPGRQREVGSLIDYIHKNDNVQRELDWGLSFDNSLSSGILQTEKHQGGKT 382  
 366 TARSAPD-KQBEISLHK--NASTNLDPIYIEFGIKVMDTENTYGVLPAPILQYGRN 422  
 363 PDY-NPQPADWSKETRGAPLISVXPLDMLIYFRNYSANSLIQ---LFCVTPAMG 437  
 423 RAIAIPNGQW--DMRGQFYNGEIKYMAIACFAPQCKREEVLYKNFTDLRKISDAG 480  
 438 MQMRK---AIMEVDDBTEAVLYRLQOKTADQIVVCLSSNRKQYDAIKYKLTDC 493  
 461 MPIQQCFCKYKQADSVENFRLKNTYSG-LDLIIVL--PKGTPIYAEVKGVDITL 538  
 494 PTPSQCVARTLQKQOTVMAIATKIALQMCNKGELWRVDP-----LKVMTVIGID 546  
 539 GWATQCVQKVVK--TSPTQISNCLKINYLG--INNIIIVPQRSVAFQOPIVIFGLAD 595  
 547 CYHMDTAG---RSLAGVVASINEGMTWFSRCIFQDQSGVLDGKVCQALAAIRANSC 603  
 596 VTHP-PADGKKRPSITAVAGSDAHPSRYCATVVRQPSRQSTIEBLSYMWELLIQYKS 654  
 604 NEYMSRIIVYDGVGQDLKTLVNYEVPQFLDLKLSIGRGYNPLTVIVYKKEVNTFF 663  
 655 TRFKETRIIFRYDGVPEGLPQILHYELLARDACTIKLEKDYQDPITIVVQKHHRLFF 714  
 664 A---QSGRLONPLPGYIVDVEVTRPEMYDFPIYSQAVRSGSVSPFTYNTYVNSGLKP 719  
 715 CADKNERIGKSNIDAGITVDNITNTHPEFDFYLSHAGIQTSRPSHYVLMNDNRFTA 774  
 720 DHIGRLTYQLCHIIYNNWPGVIRVPAFCQYAHKLAF-----LVGOSIH 761  
 775 DELQILITVGLCHTYVRCRVSISIPAPAYARLVARFARYHLVDKHDGSGSHISGQNG 834  
 762 REP 764  
 835 RDP 837



QY 755 ----LVGQSHREP 764  
DB 828 EGSVHSGQSGNRDP 841

RESULT 8  
ID 12C2\_RABIT STANDARD; PRT; 840 AA.  
AC 07/503;  
DT 16-OCT-2001 (Rel. 40, Created)  
DR 15-MAR-2004 (Rel. 43, Last sequence update)  
DE 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2) (Fragment).  
GN EIF2C2.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxId=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=98267198, PubMed=9602122;  
RA Zou C., Zhang Z., Wu S., Osterman J.C.;  
RT "Molecular cloning and characterization of a rabbit eIF2C protein.",  
RL Gene 211:187-194(1998).  
CC - FUNCTION: Plays an important role in the eukaryotic peptide chain  
initiation process.  
CC - SIMILARITY: Belongs to the argonaute family.  
CC - SIMILARITY: Contains 1 PAZ domain.  
CC - SIMILARITY: Contains 1 PAZ domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by, and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).CC  
CC EMBL AF005355; AAC24323.1; ALT\_INIT.  
DR PIR; P05505; J05669.  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003105; PAZ.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; PAZ; 1.  
DR PROSITE; PS00821; PAZ; 1.  
DR PROSITE; PS00822; PAZ; 1.  
KW Initiation factor; Protein biosynthesis.  
FT NON\_TER 1  
FT DOMAIN 216 329 PAZ.  
FT 498 799 PAZ.  
SQ SEQUENCE 840 AA; 95305 MW; 1E703F9E31391F29 CRC64;

Query Match 12.3%; Score 503.5; DB 1; Length 840;  
Best Local Similarity 24.1%; Pred. No. 2e-26;  
Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;

QY 20 GSSGIVRLSTNFRRLTSRPPALQYHIDNPLMEARLRSL---LFQR--EDLIG-K 73  
DB 13 GTSGRITKIQANFFEM-DIPKIDIIHYEEDIDPEKCPKRVNEIYEHVQHFKAQIFDR 71  
QY 74 CHAFDD-----TLFPRKLOOKVTEVFSKTRNGED--VRIT-----TLTN 113  
DB 72 KPFEDRKRLDLYTMPPLIG-REKVELLEVLPBEGKORIKKVIKXWSCVSLQALMDALSG 130  
QY 114 ELPEPTCLQFYNIIFRLIKLWNLQIGRNVYNDPIDIP-SHRLVYNGFTTSIIQ 172  
DB 131 RLPSVFEFTIQALDVVWRH-LPSMRVTPVGRSFTFASBSCSNPLGGREWFGFQSVRP 189  
QY 173 YENSINLCTDVSHKVL-RSEFTVDF---MNFYHQTSEHK-----FOEQVSKELIGLVYL 223  
DB 190 SLWKMLNIDVSAATATKQPVLEFVCEVLDFKSIEQKPLTDSQKRVKFTTEIKGLKVE 249

QY 224 TKY---NKTAYVDIDMDQNKSTFKKADG---EVSFLEYRKYVNGEITDLKQPLY 276  
DB 250 ITHGQMKRKYRVNCTVRRPASHQTFPLQOESGQVECTVAQYFQDRHCLVIRPHPLCL 309  
QY 277 -VSQPKRRRPGGTLLEGFAMLIPELCYLG-----LTKRKNDFNNKDLAVHTRLP 328  
DB 310 QVGEQKH-----TLPLEVONTIYAGRCIKKLLD---NOTSTW--IRATAFSAP 354  
QY 329 EQRQREVGKLIIDYHKNVQRELDMGKSPNSNLSTFSGRILQTEKHQGGTDYVQ 388  
DB 355 D-RQEBISKLMR--SASFITDYPVHFGIMVDEMTDVGRLQPPSIIYGR----- 404  
QY 389 FADWSKTRGAPLISYKPLDN-----MLIYTRNYEA---NSLIQNLKYT 433  
DB 405 -----NKALATPVQGVMDNRNKGPHGIEIKVMALACFAPQRCQTEVHLKSFEOIRKIS 459  
QY 434 PANGWQMRK---ALMIEVDRTYAVLVLOQKVTADTVYVCLLSNKKDKYDAIKKL 489  
DB 460 RDAGMPIQGPQPCPKYAGADSVGPMFRHL-KNTYAGQLVVVIL-PGKTPYVAEYKVG 517  
QY 490 CTDCPTPSQCVARTGKQQTWMAITKIALQNCMG-----ELMREVDIPLKLVITV 543  
DB 518 DTVLGMAITQCVQKAV--QRTTPQTLNCLKINVLGGVNNILLPQGRPVFGQVIEL 575  
QY 544 GIDCYHDMTAG--RRSIAGFVASINEGTRMPSRCIFDRQGLVDGLKVCLOALRLAW 600  
DB 576 GADVTHP-PAGDQKKSIIAAVGVSMDAHNRKCATVRYQQRHREITIQDLAMVREILIOF 634  
QY 601 NSGNEVMPRIIYRPGVGDGQKTLVYVEVQFIDCLKSIGNGVRLTVYKRVNT 660  
DB 635 YKSTRKPRIRIYRPGVSEGOQVYHHELAIRACIKLEMDYDQGLTFIVQGRHNT 694  
QY 661 RPPA---OSGRLQNPRLPGTVIVETVREYDFPVQAVRSGSVSTHNVITDNG 716  
DB 695 RLCTDKNERVSGNIPAGTVDTKITRPTEDDFYLCSHAGIQGTSRPSHVHVMDDNR 754  
QY 717 LKPDHQRLLTYKLCIHYMMWPVIRVAPCOYAKLAF 754  
DB 755 FSSDEQLITVYQLCIHYVRCITRSVSIIPAPYIAHLVAF 792

RESULT 9  
ID PINH\_ARATH STANDARD; PRT; 988 AA.  
AC Q9XGM1; O49256;  
DT 16-OCT-2001 (Rel. 40, Created)  
DR 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE PINHEAD protein (ZWILLR protein).  
GN AT5G43810 OR MOD19.17.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxId=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Landsberg erecta;  
RX MEDLINE=9094873; PubMed=9876176;  
RA Lynn K., Fernandez A., Alda M., Sedbrook J., Tasaka M., Mason P.,  
RT "The PINHEAD/ZWILLR gene acts pleiotropically in Arabidopsis  
development and has overlapping functions with the ARGONAUTE1 gene.",  
RL development 126:469-481(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Landsberg erecta;  
RX MEDLINE=96169396; PubMed=9501101;  
RA Mousavian B., Schoof H., Haeccker A., Uerogens G., Laux T.;  
RT "Role of the ZWILLR gene in the regulation of central shoot meristem  
cell fate during Arabidopsis embryogenesis.",  
RL EMO J. 17:1799-1809(1998).

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneo T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (Apr 1999) to the EMBL/GenBank/DBS databases.
CC -1- FUNCTION: Developmental protein required for reliable formation of
CC primary and axillary shoot apical meristems. May be a component of
CC a hypothetical meristem forming competence factor.
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 P1w1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF154272; AAD40098.1; -
CC EMBL: AJ223508; CA11429.1; -
CC EMBL: AB026551; BAB11310.1; -
CC PIR: T52134; T52134.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR003165; P1w1.
CC Pfam: PF02170; PAZ; 1.
CC Pfam: PF02171; P1w1; 1.
CC PROSITE: PS00821; PAZ; 1.
CC PROSITE: PS00822; P1w1; 1.
CC Developmental protein.
CC DOMAIN 337 451
CC FT DOMAIN 625 946
CC FT CONFLICT 475 479 PRDR -> ABCQR (IN REF. 2).
CC FT CONFLICT 671 671 N -> D (IN REF. 2).
CC SEQUENCE 988 AA; 110867 MW; 32EBB349C61DA20 CRC64;

Query Match 12.3%; Score 503; DB 1; Length 988;
Best Local Similarity 23.7%; Pred. No. 2,7e-28;
Matches 204; Conservative 136; Mismatches 399; Indels 122; Gaps 32;

QY 5 VNTRONLHVYKESKSGGIVRLSTNFRSLRSPQWALYOHIDYDNLMEARLRSL- 63
DB 120 MGVKNSNAPRPGGTIGTKIVKANHF-LADFTKOLNDYDVTITEVSGSKVNRALI 178
QY 64 -----LFOHEDLIGKCHAFD-----GTILFLPKLQOKYTE---VPSKTNGEDVR 106
DB 179 AELVRLYKESDGRRLPAYDGRKSLYTAGELPFTKESVYKIVBEDDIIINGPKERSYK 238
QY 107 IIT-----LTNELPTSPICLOFYNIIFRLKIKIMLQOIGNYVNPNDPI 153
DB 233 VALKVARANMHLSEFLAGKADCPQBAVQILDIVLRE-LSVKRCFVGSFSPSD-I 295
QY 154 DIP---SHRLVTPGFTTSIILOYENSIMLCYDV-SHKYLRSETVLDENFNFYHOTEEK- 208
DB 296 KIPQRLSEGLSEWCGFYQSIPTQGLSLINIDMAAFAIEPLPVEFAQLIGKDVLSKP 355
QY 209 -----FOEVSKEILGLVNLTKN---NKTYYRVDIDWQONPKSTFKADGSEV-SLEFY 260
DB 356 LSDSRVAKIKGLRGVKEVTHRANVRKRIRVAGLTPQTRFLNFPVDENCTKASVETIF 415
QY 261 RKOYNOBITDLKQPYL-VSQPRRRRPGGLTLPGPMLIPELCYLTGLTDKXRNDFNWKD 319
DB 416 GMYGFTTIOHTLPCLOYONCK-----ASYLPMEACKIVEGORYT-----KELNEKOTYAL 467
QY 320 LAVHRLRPEQORQREVRGLIDYTHKNDVQRRLRWGSLFSDNLSFGSLQTE--KIH 377
DB 468 LKV-TCQRDRDENILRTVQ--HNAYQDDPYAKKFEKNISSEKLASVEARILLPAWLKTH 524
QY 378 OGKTFDYVPOPADMSKETRGAPLISVKKPLDWMLLYTRRNYE--AANSLIQNTFKYTPA 435

```

---

```

DB 525 ENGEKDCLPQVGQMMNMKN-KIMNGMTVSRNACVFSRSVQENARFGCNELQMWCEV 562
QY 436 MGNOMRAIMIEV-----DDETE-----AYDRLVQOKVTADTQIVCLLSNRKDYDAIK 486
DB 583 SGMEFNEPEVPIPIYSARPQVEKALGHVHTSNKTKGKELELLAILPNNNSISVGDLE 642
QY 487 KYLCTDPTSPQCVAR--TLGKQOTWALATKIALQNNCKNGE-----LNRV-- 533
DB 643 RICETELGLISQCLTGHVYKIKQ-----YLANVSLKINVKNGRRTVLDIAISCRIP 697
QY 534 --DIPLLVMIYVIGDYCHMTA--GRRSIAGFVASIN-EGMTWPFRCIFQDQGOELVDG 588
DB 698 VSDIP--TIIFADYTHPENGESESSPSIAVVASQDMEVITYKAGLVCAQARQELIQD 754
QY 589 LKVCLOQALR-----AMNSCNEVMSRIIVRYDGVGDQQLKTVNYEVPQFL 635
DB 755 LYKWDQPVAGTVSGGMIRDLISFRKATGQKRLIIFRYDGVSEQGVLLYELDAIR 814
QY 636 DLKSLIGRGNPRLTYIVKKRVYTFEFA-----QSGGLQNPFGTVIDVETPEWY 689
DB 815 KACASLEPNYPVTEIIVQKRHTRLFANNHRDKNSTDSGNILPQVVDYTKIHPTEF 874
QY 690 DPEIVSQAVRSGVSPFTHVNYVINDSGLKPDHIOQLTYKCHIVYMPGVIRVAPCOYA 749
DB 875 DFYLCSHAGIGTSRPAHNVHVDENNFTADGIQSLTNMLCYTARCTSVSIVPAYYA 934
QY 750 HCLAFVGSIHREPILSLN 770
DB 935 HLAFAFR--RFYLEPEIMQDN 953

RESULT 10
ID 12C3_MOUSE STANDARD; PRT; 860 AA.
AC Q8CJF9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 3 (eIF2C 3) (eIF-2C 3)
DE (P1w1/argonaute family protein meif2C3).
GN eIF2C3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN 11]
RP SEQUENCE FROM N.A.
RA Doi N., Zemo S., Yi-Tei K., Takahashi F., Ueda R., Miyata Y.,
RA Saigo K.;
RT "eIF2C family proteins and Dicer homologs are required for siRNA
RT mediated RNAi in mammalian cells."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBS databases.
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process (by similarity).
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 P1w1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB081473; BAC15768.1; -
CC MGD: MGI:2140548; AW048686.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR003165; P1w1.
CC Pfam: PF02170; PAZ; 1.
CC Pfam: PF02171; P1w1; 1.
CC PROSITE: PS00821; PAZ; 1.

```



QY 323 HTRLRPEQOREVGLIDYIHNKDNVQRLRDMWGLSPDSNLSFSGRLQTEKHQGGKT 382  
DB 366 TRKASAD-RQBEISRLMK--NASCNDLPYIQEFGIKVODMEVGRVLPAPILQYGRN 422  
QY 383 PDY-NPOFADMSKETRGAPLISVPLDNLITYTRNPEANSLION---LFKVTAMG 437  
DB 423 RAIAFENQGVW--DMRGKFYNGIEIKVWAIACFAPKOCREBVLKFNFDLRKISKDAG 480  
QY 438 MGRK-----AIMEVDRTREALYLVLOQKVTADQIVVCLSSNRKDYDAIKKYLCTDC 493  
DB 481 MFIQGPFCCKYAGADSVEMFRLKNTYSG-LQLIIVIL--PGKTPYAEVGRVDTLL 538  
QY 494 PTPSOCCVAVRTLGKQOTWAIATKIALQNMCKMGELEMRVIDP-----LKLVMIVGID 546  
DB 539 GMAIQCVQKNAVK--TSPQTLNCLKKNVGLG--INNILVPHGRSAVFOQPVIFLGAD 595  
QY 547 CYHDMTAG--RRSIAGFYASINEGNTWFSRCIFQDRGQELVDGLKVCLOALPAMNSC 603  
DB 596 VTHP-PAGDGKPSITAAVGSMDAPSRXCATVRVQRPQOEIIEDLSYWEELLIOFYKS 654  
QY 604 NEMPSRIIVYDGVGDQGLKTLVNYEVPQFLDCLKSIGRGVNPRLTVYKKRYNTRFF 663  
DB 655 TRFKETRIIFRDGVEPQGLPOLHLELAIDACIKLEKDYQPIITIVQKRNHTLRF 714  
QY 664 A-----QSGRLQNPPLGTVIVDVEVTRPEWYDFIVSOAVRSGSVSPHTYNYLYNSGLKP 719  
DB 715 CADKXERIKSGNIDAGITVDNTNTHPEFDFYLCOSHAGIQGTSRPHSYVLLMDNRETA 774  
QY 720 DHIOGLTYLCTGIYNNWBEIVVAPCOYANHLAF-----LVGSGIH 761  
DB 775 DELQILTYOLCHTYRCTRVSIPAPAYARLVAFRARIYLVKDHSGEGSHISGSGNG 834  
QY 762 REP 764  
DB 835 RDP 837  
RESULT 12  
ID 1203 HUMAN STANDARD, PRT, 860 AA.  
AC 09H9G7; 09H1U6;  
DT 15-MAR-2004 (rel. 43, Last sequence update)  
DT 15-MAR-2004 (rel. 43, Last annotation update)  
DE Eukaryotic translation initiation factor 2C 3 (eIF2C 3) (eIF-2C 3)  
GN EIF2C3 OR AGO3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masudo Y., Oshima A.,  
RT "NEO human cDNA sequencing project."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 221-860 FROM N.A.  
RA Brown A.;  
RL Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain  
CC initiation process (by similarity).  
CC -1- SIMILARITY: Belongs to the argonaute family.  
CC -1- SIMILARITY: Contains 1 PAZ domain.  
CC -1- SIMILARITY: Contains 1 Pwi domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AK022827; BAB14262.1; -  
DR EMBL; AL138787; CAC21457.1; -  
DR Genbank; HGNC:18421; EIF2C3.  
DR MIM; 607355; -  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pwi.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pwi; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; Pwi; 1.  
KM Initiation factor; Protein biosynthesis.  
FT DOMAIN 236 PAZ.  
FT DOMAIN 518 Pwi.  
SQ SEQUENCE 860 AA; 97330 MW; BED26B529A0CF920 CRC64;  
Query Match 12.1%; Score 497; DB 1; Length 860;  
Best Local Similarity 24.3%; Pred. No. 6e-28;  
Matches 207; Conservative 150; Mismatches 354; Indels 140; Gaps 37;  
QY 20 GSSGIIVRLSTNHRFLTRPQWALYOHYHIDYNPLMEARIRLSAL---LFQHEDLI---GK 73  
DB 24 GAMGKPIKILANFCOV-EIKIIVLYEVDIKDKCPRRVRNREVDVMQVHFVYITFGDR 82  
QY 74 CHAFDG-----TLFLPKRLQGVTEFVSSTRNGED-----VRITITLTNE----- 114  
DB 83 RPYVDGRSLXYTANPLVATVATGVDLDTLPGEGKDRPFVYSIKFYSRVSWMLHLEVLTG 142  
QY 115 -LP-----PTSPITLQFYNIIFRLIKINMLQOIGNYYNPNPDIDP-SHRLVIMP 164  
DB 143 RTLPPELELDKPISTNVNVAHDVVLNR-LPSMKYTPVGRGFFSAPBEGYDHPILGGREVMF 201  
QY 165 GFTTSIIQYENSIMLCVDVSHKVL-RSEYVLDRY-----FNFYHQ-----TEBHKPQEOV 213  
DB 202 GFHQSVRPAWKKMLINDVSATGAFYKAQPIVQICMEVLDIHNTDECPRLDLSRV--KF 259  
QY 214 SKELIGLVLTXY---NNKTYRVDIDMDQNPSTF---KKADGSEVSELEYRKYQYQ 266  
DB 260 TKEIKGLKAVTHGCMTRKRCVONTRRPASHQTPFLQENQGVTERVTAQFREFKYL 319  
QY 267 EITDLKQPVL---VSQPKRRRGGTLPQPMALIPELCYLTG-----LTDKMRDYN 315  
DB 320 Q---LKYFHLPCLOVQGEQGH-----TYLPLEVCNIVAGORCIKLTLD--NOTS 363  
QY 316 VMQDLAVHTLTPROREVEGRILDIYHNKDNVQRLRDMWGLSPDSNLSFSGRLQTEKHQGGKT 373  
DB 364 TW--IKATASAD-RQBEISRLV---RSANETDEPFOGEPFKRDEMAHATGRVLP 416  
QY 374 EKHQGGKTEDY-NPOFADMSKETRGAPLISVPLDNLITYTRNPEANSLION 428  
DB 417 PMLQYGRNRTVATPSHGVW--DMRGKFYNGIEIKWAIACFAPKOCREBVLKFNFDLRKISKDAG 474  
QY 429 LFKVTAMGMGRK-----AIMEVDRTREALYLVLOQKVTADQIVVCLSSNRKDYDAIKKYLCTDC 484  
DB 475 LRKISKDAGMPLOQPCFCYKAGADSVEMFRLKNTYSG-LQLIIVIL--PGKTPYAEVGRVDTLL 532  
QY 485 IKKYLCTDCPTSPSOCCVAVRTLGKQOTWAIATKIALQNMCKMGELEMRVIDP-----LKLVMIVGID 538  
DB 533 VKRIVGDTLLGMAIQCVQKNAVK--TSPQTLNCLKKNVGLG--INNILVPHGRSAVFOQPVIFLGAD 590  
QY 539 LVNIVGIDCYHDMTAG--RRSIAGFYASINEGNTWFSRCIFQDRGQELVDGLKVCLOALPAMNSC 595  
DB 591 PVIPLGADVTHP-PAGDGKPSITAAVGSMDAPSRXCATVRVQRPQOEIIEDLSYWEELLIOFYKS 654  
QY 596 ALPAMNSCENYMSRIIVYDGVGDQGLKTLVNYEVPQFLDCLKSIGRGVNPRLTVYKKRYNTRFF 655  
DB 650 LLIQFYKSTFRKPIIFRDGVEPQGLPOLHLELAIDACIKLEKDYQPIITIVQKRNHTLRF 709



```

QY 656 KRVNTRF-----AOSGRLQNP.LPGTVIDVETREPMDEFIYSQVRSGSVSPETHNYI 711
DB 710 KRHTTRLEFCADTRERVRSGNIPAGTVDDTITHPEFDFYLSHAGIGQTSRPSHYHVL 769
QY 712 YDNGSLKPDHIOQLTYLQCHIIYNNPGVIRVAPCOYAHKLAFF-----754
DB 770 WDNCNCPADLQTLTYQLCHITVRCSTRSVISIPAPAYAHVLVPRARYHLVDKXHSAGS 829
QY 755 -LVGQSIHREP 764
DB 830 HVSQSGNSGRDP 840

RESULT 13
12C4_HUMAN
ID 12C4_HUMAN STANDARD; PRT; 861 AA.
AC Q9HCK5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Eukaryotic translation initiation factor 2C 4 (eif2c 4) (eif-2C 4)
DE (Argonaute 4).
GN EIF2C4 OR AGO4 OR KIAA1567.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res 7:273-281(2000)
CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC -1- Initiation process (by similarity).
CC -1- SIMILARITY: Belongs to the argonaute family.
CC -1- SIMILARITY: Contains 1 PAZ domain.
CC -1- SIMILARITY: Contains 1 P1WI domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AB046787; BAB13393.1; ALT_INIT.
DB Genew; HENC:18424; EIF2C4.
DB MIM; 607356;
DB InterPro; IPR003100; PAZ.
DB InterPro; IPR003165; P1WI.
DB Pfam; PF02170; PAZ; 1.
DB Pfam; PF02171; P1WI; 1.
DB PROSITE; PS50821; PAZ; 1.
DB PROSITE; PS50822; P1WI; 1.
DB Initiation factor; Protein biosynthesis.
DB DOMAIN 225 338 PAZ.
DB FT DOMAIN 509 820 P1WI.
DB SEQUENCE 861 AA; 97096 MW; F236PF05047534C1 CRC64;

Query Match 12.1%; Score 485.5; DB 1; Length 861;
Best Local Similarity 24.0%; Pred. No. 7.7e-28;
Matches 206; Conservative 149; Mismatches 353; Indels 149; Gaps 35;

QY 20 GSSGIVLSTNHFRLSRPOMALYQYHIDVYPLMEARLSALL-----FOHEDLIGK 73
DB 22 GTVGKPIRLANHFQV-QIPKIDVYHYDVDDIKPEGRPRVRVREYVDTWRFHKKQIFGDR 80
QY 74 CHAPDG-----TTLPLP---KRLQKVTVEVSKTNGEDVATITL----- 111

```

```

DB 81 QPGYDKKMYTAPLPLFGRDRVDMETL-----LPBESKDQTFKYSVQWVSJQLLLEA 136
QY 112 -----TNEIPSPICLOEYNNIFRRLKIMNLQOIGNVYVNPIDIP-SHRLVWPGF 166
DB 137 LAGHLNEVPDSDVQALD-----VITFHLPSMKYTVGASFFSPBEGYHPLDGGSEWFGF 192
QY 167 TTSILOYSNSIMLCTDVSFKVL-RSETVLDMF-----NFYQTEHKEQEQY-SKEL 217
DB 193 HQSVRPAWMMNMLNIDVGFATFRAQPIEFMCEVLIDQINNEQTKPLTDSQRYKFTXEI 252
QY 218 IGLVLTLYK---NKKTRVYDIDMDQNPSTF-----KKADSEVSFLYEYKQYNOEITD 270
DB 253 RGLKVEVTHGQMRKRVKCVNTRPASHQTFPLQLENGAMECTVAQIFKQKSLQ--- 309
QY 271 LKQVPL-----VSOPRRRRPGGTLFGPAMLIPELCTYLG-----LTDXRNDPNWMD 319
DB 310 LKYPHLPCLQYQGRQKH-----TYLPLEVQNIYAGRCIKLTD---NOTSTM-- 354
QY 320 LAVHTRLTPRQPRQREVGRLIDYHKNDVQSELEPDWGLSPNSNLTSRGRILQTEKHQ 379
DB 355 IKATARSAPD-ROEISRLVKSNSMVGPPDYLKEFGIVNEMTELTGRVLPAPMIOYG 413
QY 380 GKTPEY-NPOFADMSKERTGAPLISVKPLDNWMLITYRBYEANSILION-----LFKTP 434
DB 414 GRNKTVALPNOGW--DMRGKQFAGIEIKYMAVACFPQKQCRBDLKSFTDQLRISK 471
QY 435 AMWQMRK---AIMEVDRTFAYLVLQOKVT-ADTQIVCLISNRKQDAIKYL 489
DB 472 DAGMPDQGPCFCCKYAGADSEVEPMFKL--KMYVGLQTLIVIL-PKTPYAEVKKVG 528
QY 490 CDDCTGCPQCVARLTGQGVVMAIATKIALQNMCKMGEIMRVIDP-----LKLVM 542
DB 529 DTLGMAIQCVQKVVK--TSPQLSLCLKINKKGG-INNVLPVHQRSVPQQPYLF 585
QY 543 VGIDCYHMDTAG---RSIAGFVASINEGWTWRSCTFQDR-----GQELVUGL 589
DB 586 LGADVTHP-PAGDGGKPSIAAVGSMDSHPKRYCATVAVQTSROEISQELLYSGQVIDL 644
QY 590 KYCLQALRAWNSCNEWPSRIIVYRDVGQQLKTVLNYEVPQFLDLKISIGRGYNRL 649
DB 645 TMMVBELIQFYKSTRFPRIRITRYGVSQGMQVAMPBLIARKKACISEEDYRGI 704
QY 650 TVIVYKRVNTRF-----AOSGRLQNP.LPGTVIDVETREPMVDFYYSQVRSGSVSP 705
DB 705 TVIVQKRHTRLFCADKTERVKGSGNIPAGTVDDTITHSEFDFYLSHAGIGQTSRP 764
QY 706 TYNVTVNDSGLKPDHIOQLTYLQCHIIYNNPGVIRVAPCOYAHKLAFF-----754
DB 765 SHYQVLMDCNCPADLQTLTYQLCHITVRCSTRSVISIPAPAYAHVLVPRARYHLVDKX 824
QY 755 -----LVGQSIHREP 764
DB 825 DSAEGSHVSGQSGNDP 841

RESULT 14
AGOL_ARATH
ID AGOL_ARATH STANDARD; PRT; 997 AA.
AC Q9SJK3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute-like protein At2g27880.
GN AT2G27880 OR TIE2.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid1; Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

```

RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.B., Feldblyum T.V.,  
 RA Buehl C.R., Ketchum K.A., Lee J.J., Roming C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,  
 RA Niemann W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.,  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 CC -1- SIMILARITY: Belongs to the argonaute family.  
 CC -1- SIMILARITY: Contains 1 PAZ domain.  
 CC -1- SIMILARITY: Contains 1 P1w1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, AC006929; AAD21514.1; -.  
 DR PIR, A84678; A84678.  
 DR InterPro: IPR003100; PAZ.  
 DR InterPro: IPR003165; P1w1.  
 DR Pfam, PF02170; PAZ; 1.  
 DR Pfam, PF02171; P1w1; 1.  
 DR PROSITE, PS00821; PAZ; 1.  
 DR PROSITE, PS00822; P1W1; 1.  
 DR DOMAIN 359 471 PAZ.  
 DR FT 638 958 P1W1.  
 DR FT DOMAIN 638 958 P1W1.  
 SQ SEQUENCE 997 AA, 111134 MW, DA93A1446C422F31 CRC64;  
 Query Match 12.1%; Score 495.5; DB 1; Length 997;  
 Best Local Similarity 24.1%; Pred. No. 9,5e-28;  
 Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;  
 QY 20 GSSGIIVRLSTNHPRLTSRQWALYQYHIDVYPLMEA-----RLRSALLFQHEH--LIGK 73  
 DB 154 GILGKWKVVRANHF-LVQVADADLVHYDVSTNPEVTSKVRNWKLVKVKYKSHLGK 212  
 QY 74 CHAFD-----GTLFLPKRLQOKVTEVFSTKNGEDVRITITLNLPTPTSCIQF 125  
 DB 213 SPAYDORSLYTAGPLPFDSEKFEVNLAEKRDGSSGKDRPKVAVKN---VTSTDLTYQL 269  
 QY 126 YNIIIRRR-----LLKIMNL-----QOIGRNYNPNPDPIIDISHR-----LVI 162  
 DB 270 QOFLRKORREAPDYDTQVLDVLRDKPSNDVSVRSFPHYSLGKQARDGSGELGDIY 329  
 QY 163 MEGFTTSILQYENSIMLCTDVSHKYLRSSTVL-DFMFNFYHQTBEHK-----FOBQVSK 216  
 DB 330 WNGVGSGLRLTQMGSLINIDVARSFYEPVIVTDTFSKFLNRDLNRLPARDSDRLKVKV 389  
 QY 217 LIGL-VLTGVN-NKTYRVDVDDMDONPKSTPKKADGVSSTLEYRQYQNETIDLKOP 274  
 DB 390 LFTLTKRLAHNCTSAKISGISLPIRLRLPTLBDKSKETVQVFAEKYRNVYQALP 449  
 QY 275 VAVSQRRRGGGGLPGPAMLIPELC-----YLTGLTDRGRDFFVMKDLAVHTLTLT 327  
 DB 450 ALQI-----GSTTRVYVLPMLCCQJDEGORTKRLNEKQ-----VTALLKATQR 494  
 QY 328 PQQRQREVRLLIDYHKNQNVORELDWGLSPDSNLSGSILOTG-KIHQSGKTDY 385  
 DB 495 PPRRNSIKRLVKNVNDLSKE--FGMSVTTQLASLEAVLPPLPKYHDSSEKKNV 551  
 QY 386 NQOFNAMS-----KETGAPLIVSKPLDNLVLI--YTRRNYEAAANSLIQWLFVNTPAMQ 439  
 DB 552 NPLRGKNNVLDKKNVNGAKVTS-----WTCVSPSTRIDRLGPOFCKQLIOMCVSKMB 605  
 QY 440 MKKA-----IMTEVDRTAEVLRVLOQKVTADTQIVVCLSSNRKDKDAIKKYLCTD 494

DB 606 FKQPAIPFISCPPEHIEERLLDI--HKRAGPLQLIVL-PDVNGSYKIRICRIGTEIG 662  
 QY 495 TFSQCVAAATLQKQVMAIAIKIALQNCCKRG-----ELMRDIPL---KLWIVGI 545  
 DB 663 IVSQCCQPRQVKNL--KQYMENVALKIVKIGTGRVTLNDIAIRRIPLITDRPTIMKA 720  
 QY 546 DCHMDMTA--GRSINGFVASIN-EGMTWFSRCIPQDGOEVLDPDLKCYLQALAR----- 598  
 DB 721 DVTHPQGEDSSSLAIVASMDWPEINKRGIVSAQAREPILQILYLVQDPQKGLVH 780  
 QY 599 -----AMNSCNEYSPRIIVYDGVGQKTLVYEVYQFIDCLKSIGRGYNPL 649  
 DB 781 SGLIREHFLAFRRATQIPRIIFRYDVGSEQFQVLLHMENTAIRKACNSIQENTVPPV 840  
 QY 650 TVIVYKKRVNTRFFAASG-----RLQNPLEGTVIVDEVTPREWDPIFVSAVRSASV 703  
 DB 841 TFIIVQKRHRTRLPFQHGKRDNTDSGNIQETVVDYKICHPNEDFYLNHAGIGTS 900  
 QY 704 SPFHVVIYDNGSLKPDHILQRLTYKLCIHYNMPGVIRVAPQYAKLAF 754  
 DB 901 RPAHYHVLIDENGFTADQQLMTNNLCITYARCTKSIVIPAYYAHAAF 951  
 RESULT 15  
 ID 1202 HUMAN STANDARD, PRT, 851 AA.  
 AC Q9UKV8; Q8TCZ5; Q8WV58; Q9GID1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).  
 GN EIF2C2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21912064; PubMed=11914277;  
 RA Mourtelatos Z., Dostie J., Pauskhin S., Sharma A., Charroux B.,  
 RA Abel L., Rappalber J., Mann M., Dreyfus G.,  
 RT "tRNAs: a novel class of ribonucleoproteins containing numerous  
 RT microRNAs.";  
 RL Genes Dev. 16:720-728(2002).  
 RN [2]  
 RP SEQUENCE OF 231-851 FROM N.A.  
 RX TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Krausner R.L., Pelngold E.A., Grouse L.H., Derge J.G.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang D., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Streptelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,  
 RA Brownstein M.J., Uedini T.B., Tochiyuki S., Carninci P., Prange C.,  
 RA Bata S.S., Loquellano N.A., Peters G.J., Melek J.A., Guaranate P.H.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guaranate P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
 RA Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherren A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 267-851 FROM N.A.  
 RA Kainline N., Chen X., Rolfe A., Halleck A., Hines L., Eisenstein S.,  
 RA Koundinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin Y.,

RA Phelan M., Farmer A.;  
 RT "Cloning of human full-length cDNAs in BD Creator(TM) system donor  
 RT vector."  
 RI Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 SE SEQUENCE OF 475-851 FROM N.A.  
 RX MEDLINE=200059943; PubMed=10534406;  
 RA Koesters R., Adams V., Betts D., Moos R., Schmid M., Siemann A.,  
 RA Hassem S., Weitz S., Lichter P., Heltz P.U., von Knebel Doeberitz M.,  
 RA Birner U.;  
 RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,  
 RT genomic organization, localization to chromosomal bands 1p34-p35, and  
 RT expression."  
 RL Genomics 61:210-218(1999).  
 CC -1- FUNCTION: Plays an important role in the eukaryotic peptide chain  
 CC -1- Initiation process.  
 CC -1- SIMILARITY: Belongs to the argonaute family.  
 CC -1- SIMILARITY: Contains 1 PAZ domain.  
 CC -1- SIMILARITY: Contains 1 Pw1 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AY077717; AAL76093.1;  
 DR EMBL; BC007633; AAH07633.1; ALT INIT.  
 DR EMBL; BC018727; AAH18727.1; ALT\_INIT.  
 DR EMBL; BT007229; AAB35893.1;  
 DR EMBL; AF121255; AAF13034.2;  
 DR Genew; HGNC:3263; EIF2C2.  
 DR GO; GO:0003743; F:translation initiation factor activity; NAS.  
 DR GO; GO:0006412; P:protein biosynthesis; NAS.  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; Pw1.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; Pw1; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; Pw1; 1.  
 KM Initiation factor; Protein biosynthesis.  
 FT DOMAIN 227 340  
 FT PAZ.  
 FT DOMAIN 509 810  
 FT Pw1.  
 FT CONFLICT 556 556 C -> W (IN REF. 4).  
 FT CONFLICT 581 581 Q -> E (IN REF. 4).  
 FT CONFLICT 609 609 S -> R (IN REF. 4).  
 FT CONFLICT 629 629 E -> K (IN REF. 1).  
 SQ SEQUENCE 851 AA; 96798 MW; 31A641ABC7D1D84 CRC64;

Query Match 11.9%; Score 487; DB 1; Length 851;  
 Best Local Similarity 24.0%; Pred. No. 3,1e-27;  
 Matches 192; Conservative 146; Mismatches 341; Indels 120; Gaps 32;

QY 39 PGMALYOHYHIDYHPLMEARLRSLAL--LFOH--EDLIG-KCHAFDG-----TLPFPRK 87  
 DB 42 PKIDYHIELDIKKPKRRVRNREIVHMVQHFKTQIFGDRKPVDRKNLYTAMPPIG 101  
 QY 88 LQOKYTEVFSKTRNGED--VRIT-----TLNMLPPTSPCLQFYNILFFR 132  
 DB 102 -RDKVELEVTLPGEKSKRIKFKVSIKWSCVSLQALHDALSGRLPSVFETIQALDVMRH 160  
 QY 133 LKIKMNLQOIGRNTYNNPDIPIDIP-SRLVIWPGFTTSILOYENSIMLCTDVSHKVL-RS 190  
 DB 161 -LPSWKRYTPVGRSFTASBGCSNPLGGREYVFGFHQSVRPSLWKMLNIDVSATAFYKA 219  
 QY 191 ETVIDF--MENFHYQTEEHK-----POEYVSKELIGLVVLTXY--NNKTYRVDDIDMD 239  
 DB 220 QPVTEFVEVDLFSIEQKPLDSDQVYKFTKEIKGLKLVITHCQGMKRTKRYCANTRR 279  
 QY 240 QNPKSTFKKADG---EVSLEYRKQYNQETIDLKQPVV--VSQPKRRRGPGGTLPGPA 294

DB 280 PASHQTFPLQOESQVETVAQYFKDRHKLVIRYPHLPCLQVQEOQH-----T 329  
 QY 295 MLIPELCYLTG-----LTDKMRDNFVMDLAVHTRLPEDQROREVGRLIDYHKNDN 347  
 DB 330 YLPLEVCNIIVAGQRCIKRLTD---NQSTW--IKATRSAPD-RQEISKLMR--SASFN 381  
 QY 348 VQREIRDMGLSPDSNLSFSGRILLQTEKIHOGKTFPDYNPQFADMSKETRAPLISYKPL 407  
 DB 382 TDPVYREFGIWVKDEMDDVTGRVLPQPSILYGR-----NKALATFVQGVWDM 429  
 QY 408 DN-----MLLYTRRNYEA---NSLIQNLFKYTPAMQMRK---AIMIEV 448  
 DB 430 RNKQPHTEIKKWAIAICFAPQROCTEVLHLSFTEQRLKSRDAGWPIQGPFCCKVAQG 489  
 QY 449 DDRTEAYLRVLIQKVTADTQIVCLSSNRKDKYDAIKYLCIDCPFPBSQCVVARTIGKQ 508  
 DB 490 ADSVEMFRHH-KMYAGLQLVVIL-FGKTPVYAEVRVGDVIGMATQCVQKMY--Q 545  
 QY 509 QTVMAIATKIALQNNCKWGG-----ELMRVDIPLKLVIVIGIDCYHDMTAG--RRSIA 559  
 DB 546 RTTPQTLNLCIKIVKLGANNILLPQGRPPVQGPVIFLGADVTHP-PAGDCKPSIA 604  
 QY 560 GFVASINQMTWPSRCIFDQGGELYDGLKVCLOALPANNSCNEYPSRIIYRQVG 619  
 DB 605 AVVGSMDAHPRYCATVAVQHRQEIIDLAAVRELLICFKSTRFETRIIFYRQVGS 664  
 QY 620 DQGLKIVNYVYVPOELDKSIGGYNRLVIVYKRVNTRFPA---QSGRLQWPLP 675  
 DB 665 EGQFOQVHHEHLAIREACIKLEKYOQGITFIYQKSHHRLPCTDKNRNVGSGNIPA 724  
 QY 676 GTVIDEYTRBEMWDFPIVSAVRSVSPTHYNYIYDNGSLKPDHIIQLRTYKLCIHYN 735  
 DB 725 GTTVDTKITHPEFDYFLCSHAGIQGTRSPSHYVLMWDRNFSDBELQILTYQLCHTVR 784  
 QY 736 WFGVTVRVPAPQYAKLAF 754  
 DB 785 CTRSVSIPAPAYYALVAF 803

Search completed: May 5, 2004, 15:13:09  
 Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:10:30 ; Search time 20 Seconds

(without alignments)  
3727.422 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVTNRQNDHVESKTKG.....VGOSIHREPNLSLNRLYYL 775

Scoring table: BLOSUM62

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1295.5	31.6	824	2	T20351
2	1162	28.3	722	2	T30995
3	557.5	13.6	834	2	T41568
4	508.5	12.4	1014	2	H86438
5	505.5	12.3	891	2	T32079
6	498.5	12.2	1000	2	T23391
7	493.5	12.1	997	2	A84678
8	492	12.0	988	2	T52134
9	488.5	11.9	930	2	A84668
10	482	11.6	813	2	JC6569
11	434	10.8	887	2	T01113
12	429.5	10.5	1032	2	T23164
13	429.5	10.5	1035	2	T23165
14	421	10.3	958	2	S41013
15	421	10.3	1040	2	D88568
16	421	10.3	998	2	T33175
17	343.5	8.4	1139	2	T33175
18	316.5	7.7	905	2	T15195
19	283	6.9	892	2	T23965
20	280.5	6.8	945	2	T22933
21	253	6.2	965	2	T16676
22	252	6.1	1057	2	T19324
23	248.5	6.1	1032	2	T16114
24	240	5.9	1030	2	F88925
25	238	5.8	990	2	T15179
26	230.5	5.6	697	2	T18974
27	225	5.5	939	2	G87774
28	223	5.4	324	2	W31913

30	219.5	5.4	674	2	T19268	hypothetical prote
31	219	5.3	363	2	T34486	hypothetical prote
32	200	4.9	567	2	T34439	hypothetical prote
33	189.5	4.6	706	2	H70425	conserved hypothet
34	168	4.1	944	2	T19911	hypothetical prote
35	145	3.5	713	2	H84464	hypothetical prote
36	129	3.1	185	2	T27784	hypothetical prote
37	129	3.1	709	2	G96610	probable disease r
38	123.5	3.0	2035	2	AC0233	yeastinabactin bio
39	123.5	3.0	2041	2	T17439	peptide synthetase
40	120	2.9	1261	2	S75130	sensory transducti
41	118	2.9	1120	2	H71664	transcription-repa
42	114.5	2.8	474	2	A64691	type III restricti
43	114	2.8	2035	2	A48654	probable polyketid
44	113.5	2.8	560	2	C38604	poly(3-hydroxyalka
45	113	2.8	949	2	G84486	probable plasma me

## ALIGNMENTS

## RESULT 1

T20351  
hypothetical protein D2030.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T20351

R/Wilkinson, J.

submitted to the EMBL Data Library, June 1996

A/Accession: T20351

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-824 <WTL>

A/Cross-references: EMBL:273906; PIDN:CAA98113.1; GSPDB:GN00019; GSP:2030.6

A/Experimental source: clone D2030

C/Genetics:

A/Map position: 1

A/Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3

Query Match 31.6%; Score 1295.5; DB 2; Length 824;

Best Local Similarity 34.9%; Prod. No. 6.6e-86;

Matches 278; Conservative 169; Mismatches 302; Indels 47; Gaps 15;

QY	16	BSKTGSSGIIIVRLSTNFR--TSRPMALQYHIDVPLMEARLRGALFQHE--DLIG 72	
DB	40	QSKTSSGQPKCPANPIEMTOSDYIYQHYVEPTVDSKANREKLRNNVDEIG 99	
QY	73	KCHAFDGTILFLPRLQQYVEFSKTRNGEDVRLITLTLNELPTSGTCLQFYNIIFR 132	
DB	100	HHFVFDGMLLYLKHEEMQNMIEVQHPIDRLICIFKQNRFLVDPPTINIFNTIR 159	
QY	133	LKTKMNOQGRNYYNNDPIDPSHLYVMPGTTSLQYENSIMCTGVSHVLSSET 192	
DB	160	SPDALQTLQGRNENFGDSRAVPDYMGLPGETIRAYENEMLCVENRFQVAREE 219	
QY	193	VLDPMFNFYH-----QTEHKKPEQVSKELIGLVLTXYNNKTYRVDIDQPKSTF 246	
DB	220	---MYLIFKELKRSQNNQORVQKKNEMVGGTITIRYNNKLRFTYRDIYSISPSSEF 275	
QY	247	KKADGSEVSFLFYRKYQNGEITDLPVLYSQPKRKGCGTLPGSAMLPELCYTLGL 306	
DB	276	VR--DGSITLTKYKQNGYIEITVDQPTIISSEKPKQ--PGEPPQVSYIVPELCFTGL 332	
QY	307	TDKRNDFNWKADLAVTRLTLPQQRREVRGLIDYIRKNDNVQRELRDWSGPSDSNLSF 366	
DB	333	TDEKRDQFQMKELAKKTRMSPOQRLVESRLYDLSKNEKVMGCFKMGISLGDLANV 392	
QY	367	SGRILOTEKIHGGKGTIDYNQPADMSKETRGAPLIVKPLDWLLTYRANVA--ANS 424	
DB	393	QAVLKSEPL--QGEKKT--YEGKQAEWARGVECGIYRGSNMWTWIVIGPSGNSGLLSOK 449	

```

QY 425 LIQNLFKVTTPAMGKQMRKAIMIEV-----DPRTEAYLAVLQOKTADTQIVCLLSSNRK 479
DB 450 FIEEARRLGKILQVLTGSPMCVPIRGISPNIDYLEGVKGAIQVQVGEIHMLVWLADDNK 509
QY 480 DKYDAIKKYVLTCDTSPGQCVVARTL-----GKQQTVMALATKIALQNNCKKGGELMR 532
DB 510 TRYDSIKKFLVCEGCIPIQOCVNLKTLAKSGDGENKRLSVLKIYQMLCKTGALMK 569
QY 533 VDIPLKLVYIGIDCYHDMTGRSIAQFVASINEGTRMFSRCIFODRGOGLVDGLKVC 592
DB 570 VNIPKNTMIVGYDYLDYHSTLKGTGACVSTSTSDPFOFYSQTRPHENPTOLGNLTHF 629
QY 593 LQALRAMNSCNEY-MPSRIITYRDPVGDGOLKTLVNEVPOFLDCLKSIR----- 643
DB 630 VRKALKOYDSDQTLPSRLIYTRDAGDQIPIYKTEVCLVNDACDAVTDKAAELSNK 689
QY 644 -GYNRLTVIVYVKKRVNTRFPAQSGRLQ---NPLPGTVIDVEVTRPEWYDFIVSOAVR 699
DB 690 VOEKIKLAFIIVTKEVNRRILKQ-GSSIDNAINPQPGTVVDTVTRPERMDFYLVQFVN 748
QY 700 SSVSPHTNNVYDNSGLKPDHIOQLTYKLCIYNNMPGVLRVAPCOYAHKLAFLVQGS 759
DB 749 QGTVPVSNIHDDTDIGPDGQQLAFKLCILYNNMGTVRVAPCOYAHKLAFLVQGS 808
QY 760 IHRBNLSLSNRLYYL 775
DB 809 LHDANGCLRDGLFPL 824

```

RESULT 2  
 T30955  
 Hypochemical protein C01G5.2 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 29-Oct-1999  
 R/Bradshaw, H.; Steillyes, L.  
 Submitted to the EMBL Data Library, August 1999  
 A/Description: The sequence of C. elegans cosmid C01G5.  
 A/Reference number: Z20956  
 A/Accession: T30955  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-722 <BRA>  
 A/Cross-References: EMBL:U50068, PIDN:AAB37734.1  
 A/Experimental source: strain Bristol N2  
 C/Genetics:  
 A/Map position: IV  
 A/Intons: 58/3, 121/3, 202/3, 350/1, 471/3, 530/1, 609/2, 671/3  
 A/Note: C01G5.2

Query Match 28.3%; Score 1162; DB 2; Length 722;  
 Best Local Similarity 33.6%; Pred. No. 2.7e-76;  
 Matches 256; Conservative 154; Mismatches 279; Indels 74; Gaps 14;

```

QY 41 WALYQYHIDYNDPLMARRLSALLFOHEDL--IGKCHAFDGTILFELPKLQOKTEVFSK 98
DB 6 YSIYQYVPEPTVDSKATREMLAQPSVYVEIGHFVFDMLILYKEMQNQMIVQH 65
QY 99 TRNGEDVAITTLTINELPPTSCLOFYNIIFRLLKTMNQIGIRNYNPNNDIDIPSH 158
DB 66 PNDNSLICIRPKTKRFLVDDPQITINIFNTIIRSFDMKLTQIGRNYFD----- 115
QY 159 RLVMPGFTTSLQYENSIMLCTDVSHVLSSETLDDPMFNYQGTESHKQOVSXELI 218
DB 116 -----WNSBAL-----RKELNS-----CONNRQKQVQSCNEYG 145
QY 219 GLVLTJKNKTKYRVDDIDWDONPKSTFKADGSEVSLEYRKOYNQETIDLKQPLVS 278
DB 146 GSTIITRYNNKILHRTFLDNEITPLSKFOK-DGEOIILKEVFKQYDIDIDDEQFIIS 204
QY 279 QPKRRRGPGGLRGAMILPELCYLTGLTDKRNENFVMKDLAVHTRTLTPEQORQVRL 338
DB 205 ESKPKQ--PGEPPQYVIVPELCTFTGLTDEKRDFOKMKIANTHTSSPOQRDDETRKL 262

```

```

QY 339 IDYIHKNDVQRELRDWSLSPDSNLISFSGRLIQTEKHQGGKTDYNDQPADWSKETRG 398
DB 263 ITYLSQNTMMECFQWGISLQODLANVQARVLKSEPL-QGR--QYEGKQEMARGVE 319
QY 399 APLISVKPLDNNLIYTRNRYE--ANSLIQVLFKYTPAMGQMRKAIMIEV-----DDR 451
DB 320 CGIYRSNNMTMIVIPGSGNSGLAQKTAARNLGRLOVQDLEPMCKVKNIGSPNY 379
QY 452 TEAYLVLQOKTADTQIVCLLSSNRKQYDAIKKYLTCDTPPSQCVVARTL----- 505
DB 380 LELGKAIKSVGEEIHMVWLADDNKTRYSLKCYLCECPPIQOCVNLKTLGKSD 439
QY 506 -GKQQTVMALATKIALQNNCKKGGELMRVDIPLKLVYIGIDCYHDMTGRSIAQFVAS 564
DB 440 GGENKRLGSIYLVKIYQMLCKTGALMKVNTLPLKSTMIVGYLDYHSTLKGTGACVST 499
QY 565 INEGMTRMFSRCIFODRGOGLVDGLKVCLOALRAMNSCNE-YMPSRIITYRDPVGDGOL 623
DB 500 TSNDFQFYSQTRPHENPTOLGNLTHFYRKSLKQYDNNDKTLPSRLIYTRDAGDQI 559
QY 624 KILVTEVPOFLDCLKSIR-----GYNRLTVIVYVKKRVNTRFPAQ--SGGRLQ 672
DB 560 PIYKTEVCLVNDACDAVTDKAAELSNKVQEKIKLAFIIVTRVNNRILKQSSSKSAIN 619
QY 673 PLPGTVIDVEVTRPEWYDFIVSOAVRSGVSPHTNNVYDNSGLKPDHIOQLTYKLCI 732
DB 620 PQGTIVDVTVTRPERMDFYLVQFVNGQTVTPVSNIHDDTDIGPDGQQLAFKLCIL 679
QY 733 YNNMPGVTRVAPCOYAHKLAFLVQGSIRRENTLSNRLYYL 775
DB 680 YNNMGTVRVAPCOYAHKLAFLVQGSLSHDANGYLRDGLFPL 722

```

RESULT 3  
 T41568  
 argonaute-like protein - fission yeast (*Schizosaccharomyces pombe*)  
 C/Species: *Schizosaccharomyces pombe*  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C/Accession: T41568  
 R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
 Submitted to the EMBL Data Library, May 1998  
 A/Reference number: Z21991  
 A/Accession: T41568  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-834 <MOO>  
 A/Cross-References: EMBL:AL023705, PIDN:CAA19275.1; GSPDB:GN00068; SPDB:SPCC736.11  
 A/Experimental source: strain 97zh-; cosmid c736  
 C/Genetics:  
 A/Gene: SPDB:SPCC736.11  
 A/Map position: 3  
 A/Intons: 43/3  
 C/Superfamily: rabbit translation initiation factor eIF-2C

Query Match 13.6%; Score 557.5; DB 2; Length 834;  
 Best Local Similarity 22.9%; Pred. No. 3.1e-32;  
 Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

```

QY 20 GSGGIIVRLSTYHFRITSPQALVQYHIDYNDPLMARRLSALLFOHEDL----- 70
DB 16 GGLGKQITLKANFFQIISLPNETIQYHIVDGSVRPKQQLIWSKQYRGSSWY 75
QY 71 -----IGKCHAFDGTILF-----LPKRLQ--OKTEV-----FSKTRNGED 104
DB 76 NSVYDGRSMCKSGDLADGTIKNIGSESHPEISIOKSKINLHTLSQVNSKSYSD 135
QY 105 VRTITLTINELPPTSCLOFYNIIFR-----LKLKMLQOIGRNYNPNNDIDIPSH 159
DB 136 -----FOVLSSIMFDLILKPKRPSSTLFGFMISFTGNGSVLGGV----- 178
QY 160 LVIMPFTTSLQYENSIMLCTDVSHKVL-RSETVLDPM----- 197

```

Db 179 --AWGFGVQSRPNQFMSVAVNDISSAFWRNDLSLQTLMEYTDSCNVRLDTRPDLKRLS 236  
 198 --FNFYHOTEHKEHKEQVSKELIGLVLTNNKTYRVDIDWQNPKESTF-KKADGSE- 253  
 Db 237 RKFRLKATYCOHRNN-----VGTDLANRYVSIEGSSKASDPSFVRLNGEEQ 285  
 Qy 254 -VSFLFYRKQYNOGIIITLQKQPVLYSQRRRGGPGTLPGAMLIPELCYLT---GLTDK 309  
 Db 286 KISVAEYFLENHNNVRLQYPNLPCTII-----VANGAMLPYEFCEVAVKQRTAK 333  
 Qy 310 MENDFNWMDLAVHTRLPQKQREVEGRILDIYHND-NVQRELRDWGLSPDNLSPSG 368  
 Db 334 LNSD-----QTNAMIRFVAVQRPFRVQIIDDFVHQMDWDTDPYLTQYGMKIQKQMLEVPA 388  
 Qy 369 RILQTEKHOGKTDYVNPQPADMSKETRGAPLISV--KPLDNTLLI---YTRR-NYEAA 422  
 Db 389 RYLEPFSIRYGGDCTE-RPVSGRNN-LRGRFLDPFPAPFIRSMVAVKCFSTIRLPMKGI 445  
 Qy 423 NSLIQNLKFTVPAMQO--MRKAIMIEYDDR--TEAVLRVLQOKV---TADTOIVVCLL 474  
 Db 446 ENFLQTYVQTLTSLGINFMVKKPVLVADINGSVELCITLYKKAQYGNAPDYLPEIL 505  
 Qy 475 SSNRKQTDALIKYLCCTDPPSQCVARTGKQQTNAITKIALQWCKMG---ELW 531  
 Db 506 DKNSPEPYSIKRVGNITMLGVPSQCAISKHI--LQSKPOYCANLGMKINVKVGINGSLI 563  
 Qy 532 RVDILKLV--MIVGIDCH--DMTAGRSIAGFVASINSGMTBMFSRCIFDRQGLVYG 588  
 Db 564 PKSNPLGNVPTLLGQDYHFGVATGVSIAVSVDLNGCKTAYVRSQPRHDEVIEG 623  
 Qy 589 LKVCLOALRAMNSCNEYMPSRIIYRDGVGGQKTLVYEVPPFLDCLKSIGGYNR 648  
 Db 624 MDIIVVLLQGRATKQOPRIIYFRDQTSQGLSVINDELQIKKACSLSPKYNPK 683  
 Qy 649 LTVIYVKKRVNTRFPAQS---GGRLONPFGTVIDEYTRPEMNFYVQAGNSGSVSP 705  
 Db 684 IIVCTQKKHARFPTKNSDSDNGNPLPGTILKHYHVFYQDYFIIHSPSLQGVSV 743  
 Qy 706 THYNYIYDNGSLKPHIQTLTYKLCHIIYMPGVIRVAPCOYAKLAFL----- 755  
 Db 744 VHYTTLHBDICMPDQFQTLCTNLCTVYARATSAVSLVPVYVAVLVENLAKRYQDVTAD 803  
 Qy 756 -----VGQSHREPNVLSNKL 772  
 Db 804 TEVETSEASMDQEV--KPLIALSLKL 827

RESULT 4  
 H86438  
 protein T19E23.7 [imported] - Arabidopsis thaliana  
 C1Species: Arabidopsis thaliana (mouse-ear cress)  
 C1Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C1Accession: H86438  
 R1Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Matzila, I.  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A1Author: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A1Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A1Reference number: A86141; MIMD:21016719; PMID:11130712  
 A1Accession: H86438  
 A1Status: preliminary  
 A1Molecule type: DNA  
 A1Residues: 1-1014 <STO>  
 C1Cross-references: GB:AE005172; NID:g6692120; PIDN:AAF4585.1; GSPDB:GN00141  
 C1Genetic: 19E23.7  
 A1Gene: T19E23.7  
 A1Map position: 1

Query Match 12.4%; Score 508.5; DB 2; Length 1014;  
 Best Local Similarity 23.9%; Pred. No. 1.5e-28;  
 Matches 209; Conservative 151; Mismatches 345; Indels 171; Gaps 40;

Qy 5 VNTKQMLDHYKESKTGSSGII-----VRLSTNFRILTRSPQMALYOHYIDN---PLMEAR 57  
 Db 153 VATSDEKPEKRPDRG--GVAVRNVLYNHKVFNEESVIRHYDVEIKGEIPTRKVS 210  
 Qy 58 RLRSALLPQH-----EDLIGKCHARD-----TILFLKRLQOKTEVFSKTRNGEDVR 106  
 Db 211 RFLANVRQKFTVNDDEPFLPANTAYDQGNIFSAVELPTG-SYKYEYKTEMRGRST 269  
 Qy 107 ITTTLNLEPPTSPTLQFTNIIIRLLKIMLQO--IGRNYNPNDP-----DIP 156  
 Db 270 FTIKQVN-----VLKIGDLKEYMTGRSSFNPRDVLQGMVVMKEHP 310  
 Qy 157 S-----HRLVMPGFSTSLQYENSIMLCTDVSHTKLR--SEYV 193  
 Db 311 SKCMITVGSFETRETPDEDFRFGVIAKGRHILKPTAQSLSLCLDISVLAFKAMSV 370  
 Qy 194 LDMFNFYHOTEHKEHKEQ-VSKELIGLVY-----LTNNKTYRVDID 237  
 Db 371 IEYLKLYFNMSDMRQRRRDVEEELIGKVTNHRKQKGLTIVGLSMQTRKIDKFLDID 430  
 Qy 238 WDQN--PKSTFKKADGSEVSFLFYRKQYNOGIIITLQKQPVLYSQRRRGGPGTLGPRAM 295  
 Db 431 QEGNEPPEPKT-----SIVEYFRIKYGRHIVHKDIPCLDKNGRQO----- 471  
 Qy 296 LIP-ELCYTLGLDKKRNDFNVKDLAVHTR-----LTPEQKQREVEGRILDIYHNDNVQR 350  
 Db 472 FVMEFDDL--VGQIYPPKDLKDSALMLKLSLVNPOGRQGNIDMKR-ARNGSGGE 528  
 Qy 351 ELRDWGLSPDNLSPSGRILOTE--KHQGGTFDYNP--OPADMSKETRGAPLISVK 405  
 Db 529 IIGNFGKVDNTMTPEAGRLKAPSLKLAERGRVVEEPNPNQNMNLKKGVTRESI- 587  
 Qy 406 PLDNTLLI---YRRRYEANSILQNLKFTVPAMQOMRAIMIEVD-----DTEA 454  
 Db 588 -VHMAVLDFTASERFNKMPNDVNDLIDRCMLQGMQEMAPYKYSRMTLSNGAIEE 646  
 Qy 455 YLRVLQOKV-----ADTOIVCLLSSNRKDYAKIKYLCCTDPPSQCVVA--RTLQK 507  
 Db 647 LRSVIDEASRKHGARGAPTLVLCAM--SRKDDGKTLKMIETKLGVTQCFLTGPATKGG 705  
 Qy 508 QQYVMAIATKIALQWCKMG---ELKRYDIPKLK---VNIIGIDCHMDTAGR--SIA 559  
 Db 706 DQ-----YRANIALKMAKAVGSSNVELDFTSFPRKEDVMTFQADVNHPRARDMSISIV 761  
 Qy 560 GFVASIN-EGMTWFSRCIFQDRQGLVNDGL-KVCIQALRAMNSCNEYMPSRIIYVRDG 617  
 Db 762 AVGTGLAMPANRYAAVIAVQHPHKEILQFPGDACE-LYKAVVQATGKAPNKIVIPRDG 820  
 Qy 618 VGDQGLKTVNYEVPQFLDCLKSIGR-GYNRLTVIYVKKRVNTRFPA---QGGGRQNP 673  
 Db 821 VSDAQPFMVNLVE---LDVKLTFEKNGXNPKTIVIAQGRHDTREFPATNNDQSDKGVN 877  
 Qy 674 LPGVVIDEYTRPEMNFYVQAGNSGSVPTHYNYIYDNGSLKPHIQTLTYKLCHY 733  
 Db 878 PSGTIVVTKYHHPYEFYVLCSHHGIGTGKPHHYTLMBELFTSDQVQKLFEMQFTF 937  
 Qy 734 YNMPGVIRVAPCOYAKLAFLVQSGIH-----RBN 765  
 Db 938 TRCTKPVSLVPVYVADMVAF--GRMYTHASSRBN 972

RESULT 5  
 T32079  
 hypothetical protein T07D3.7 - Caenorhabditis elegans  
 C1Species: Caenorhabditis elegans  
 C1Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C1Accession: T32079  
 R1Author: B.; Wamsley, P.; Tyman, B.  
 submitted to the EMBL Data Library, July 1997



A:Description: The sequence of C. elegans cosmid T07D3.

A:Reference number: Z21121

A:Accession: T32079

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-891 <L&M>

A:Cross-references: EMBL:AF016682; PIDN:AB66187.1; GSPDB:GN00020; CESP:T07D3.7

A:Experimental source: strain Bristol N2, clone T07D3

A:Genetic:

A:Gene: CESP:T07D3.7

A:Map position: 2

A:Insertions: 34/2; 206/1; 560/1; 879/3

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 26-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

```

QY 20 GSSGIIVRLSTNHEFLTSRPMALYQHIDYDNPMEARLRISALLFOHEDLGKCHAFDG 79
DB 52 GVEGRSILRLANHPAVRI-PCGSVOHYQIDVFPPKCPRRVRYEYI-----G 96
QY 80 TILFLPKLQOKTEVSKTRN-----GSDVRLITLTNE-----LPPTS 119
DB 97 CLISSFSKQYFTNIRPVYDGKNNMTREPLIGTEPMNFVYLPQDSAVKRSVTKMKG 156
QY 120 PTCL-----QFYNIIFRLIKINLQOIGRYNPNPDIDPSHRL 160
DB 157 QVCLSLADDAMEGRVQRVHPHVAQSIDVILNH-LPSLKYTPVGRSFTPPGVK-PQMOM 214
QY 161 -----VIMGFTTSIIQYENSIMLCTDVSHKYL-RSETVLDNMFNYH----- 202
DB 215 HQESKLGGRVWGFHQSVAPRSQMKMLNDVSAATPYAMPFIEFVALELPVQALA 274
QY 203 -----QTEBEKFOEVSKEILGLVVLTKY---NNKTYRVDDIDMDQNPSTF---KKA 249
DB 275 ERRLSLDAQRVYF---TKERGLKIEITHGAVRRKYRVGNVTRRPAQOTFPLQLETG 330
QY 250 DGESEVSELEYRKQYNOETIDLKQVTL---VSQPKRRGGCGTLPGPAMLIPELC--- 301
DB 331 QTIETVAKYFFPDYK---IQLKYPHLPCLOVGEQK-----TYLPPEVCDIVP 377
QY 302 ---YLGLTDMGRDPMVKDLAVHTRLPPEOROREVRLIDYIHKDNYOREL----- 352
DB 378 GORCLAKLTDVQSTW---IKATAPAPF-REREICKVLS-----KALSLADPFA 423
QY 353 RDMGLSPDSNLLSPSGRILOTEKIHQGG-KTFDYNPQPADMSKETGAPLISVPLDNW 410
DB 424 HEFGITINPAMTEVAGRVLSAPKLLYGGHRATATAPQGVW---DMGKQPHITMEVRYT 481
QY 411 -LLIYTRNRYEANSI---IONLPKVTBAMGMQ-----RKAIMIEVDRTAVYRLV 459
DB 482 ALACGAGSHVXENDLRMTTQLOKISTDAQMPILIGPMFKYASGVH---QVDPMKYL 538
QY 460 QOKTADPQIVVCLLSNRKDYDAIKKYLCTDCTPSQCVAASTLQKQVTMAATKIA 519
DB 539 KQTVSAIQIOLIVVL--PKETPIYAEVKVSDPTVGIATQVOAK-NAIRTPPTQLSNLC 594
QY 520 LQANCKMG-----ELMRVDIPLKLVITVIGIDCHDMTAGRR--SIAGFASINEMGR 571
DB 595 LKNMVKLGAVNSILLPNVRPIENPFIPLGODITHPAAGDTRKPSIAAVGSDADPSR 654
QY 572 WFSRCIFDQGEELVGLKVCLOALRAMNSCNEYMPRIIVYRDGVGQGLKTLVNYEV 631
DB 655 YAAITRVQGHQGEIITDLYVRELVOFYRNTREKPARIVYVRGVEGQLFVAVLOVEL 714
QY 632 POFPLDLKLSIGKYNPRILTVYVKRNVTRPF---AOSGRLQNPPLPGTVIDVEMRPE 687
DB 715 RAIRACVWLSESGYQGITFIAVQGRHTRLPADKAOVGAGFIPGRTIVDVGITHFT 774
QY 688 WYDFPIVSOAVSAGVSFTHYVNVYDNSGLKPDHIOQLTLTKLCHLYNMPGVIVPAPCQ 747
DB 775 EIDFELCHAGIQGSRPSHNVMLDDMDLTADDELQQLTYQMCHTYVACTSVSLPAPAY 834

```

QY 748 YAHKLAF 754  
DB 835 YAHLYAF 841

# RESULT 6

T22391

hypothetical protein F48F7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T22391; T22324

R:Colles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19559

A:Accession: T22391

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <MTL>

A:Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone F48F7

R:Colles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19712

A:Accession: T22324

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <MT2>

A:Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone K02B9

C:Genetic:

A:Gene: CESP:F48F7.1

A:Map position: X

A:Insertions: 70/3; 128/2; 185/2; 673/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.9e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

```

QY 20 GSSGIIVRLSTNHEFLTSRPMALYQHIDYDNPMEARLRISALLFOHEDLGKCHAFDG 79
DB 145 GVEGRSILRLANHPAVRI-PCGTIHYQVDVTP-----DK----- 179
QY 80 TILFLPKLQOKY-----TEVFSKTRNGEDVRIITLTNELP-----PTSP-- 120
DB 180 -----PRVNRKILISGLISAFSKYFTNIRPVYDGKNNYTRPLPIGRBMDFDVTLPGD 234
QY 121 -----TCLQFNITIRLLIKINLQOIGN 145
DB 235 SAVERQFSVLKMWGVQVSLTLEDAMEGRVRCVPEEAVQAMDVILNH-LPSLKYTPVGRS 293
QY 146 YNNPNPDIDPSHRLV-----IMPFTTSIIQY 173
DB 294 FSPS--FVFNASGVWAGSCPPQASGAVAGASAGYTHAESTLGGREVMFQHSVRS 351
QY 174 ENSIMLCTDVSHKYL-RSETVLDNMFNYH-----QTEBEKFOEVSKEILG 219
DB 352 QMKMLNDIVSATAFYRSPVIEFLAVLELPVQALAEARRALSDAQRVYF---TKERNG 407
QY 220 LVVLTKY---NNKTYRVDDIDMDQNPSTF---KKADGESEVSELEYRKQYNOETIDLK 272
DB 408 LKEITHGQMKRKIVNCVTRRPAQOTFPLQLEFGQITECTVAKYFYDKR---IQLK 464
QY 273 QPVL---VSQPKRRRPGGTLPGPAMLIPELCYLT-----GLTDKRRDPMVKDLA 321
DB 465 YPHLPCLQVQGEQKH-----TYLPPEVCNIVPGRCIKKLTVDQSTW---IK 509
QY 322 VHTRLPPEOROREVGLI-----DYHKDNDVQRELDMGLSPDSNLLSPSGRILO 373
DB 510 ATRKAPF-REREISNLVKAERSADPFAH-----EFGITINPAMTEVKGVLISA 558
QY 374 EKIHQGKLT-FDYNPQPADMSKETRGAPLISVYKPLDNW-LLIYTRNRYEANSI---ION 428

```

```

Db 559 PKLYGKTRATYALPNOGVW--DMEGKQFHTGIDIVRWALACPAQOQKHENDIMETNQ 616
Qy 429 LEKTPANGKQK-----RKAIMIEVDDRTYAVLRVLOQKVTADTQIVYCLLSNRKDK 481
Db 617 LQR:SNAGMFIQGNPCCKYAVGE---QVEPMFKILKQNGSG-IGLVVVL-PGKTPV 671
Qy 482 YDAIKYICTCPPTPSQCVAVARTLQKQTVMAIKTALQNNCKMG-----ELMRVDI 535
Db 672 YAEVRKVDITVIGLATQCQAK--NAITTPQTLISNCLKNVGLGNGNILLPNVPRI 729
Qy 536 PLKLMIVGIDCYHDMTAGRR--SIAGFVASINEGTRWFSRCIFODRGQELVDGLKYL 593
Db 730 FNEPIIFPGCDITHPADSRKPSIAAVGSMDAHPSTYATVAVQCHRGESIDLYYV 789
Qy 594 QALTAWNSCNEWPSRIIYVRDVGQDLKTLVNEVVPQFLDCKSIGRGYNRLTVIV 653
Db 790 RELVQFRNTRFKPARIVVYRDGVEGQFNVLIQYELRAIRACMLERGYQGFIFIA 849
Qy 654 VKKRNTRFEA-----QSGGRQLNPLPGTVIDVETVRPEWDFEIVSQAVRSGSVPTHYN 709
Db 850 VQKRHTLPAVDKQDVQKAVNIPPGITVDVGIHTEFDFYICSHAGIGTISRFSHY 909
Qy 710 VIYDNGSLKPDHIOQLTYKLCHIIYNNPVGIVRVPACQYAHKLA 754
Db 910 VLMDNNMLTADQLQTLQMGCHTVYRCSRVSIPAPAYALVAF 954

```

## RESULT 7

```

A:Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84678
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.B.; Unayam, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617137
A:Accession: A84678
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-997 <STO>
A:Cross-references: GB:AE002093; NID:g4510428; PIDN:AD21514.1; GSPDB:GN00139
C:Genetics:
A:Gene position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C

```

Query Match 12.1%; Score 495.5; DB 2; Length 997;

Best Local Similarity 24.1%; Pred. No. 1.3e-27;

Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;

20 GSSGIIVLSTNHFLTSRPMALQYHIDVNEPMEA-----RLRSALLFQHEP-LIGK 73

154 GTLGKVMVNRHF-LVQVADRDLHYDVSINFEVISTVNNVKKLVKNYKDSHLAGK 212

74 CHAPD-----GTLFLPKRLQOKVTEFESKTRNGEDVTRITLTNELPPTSPCLQF 125

213 SPAYGRKSLYAGLPFDSKEFVNVLAKRADSGSKDRPKVAVK---VTSTDLQQL 269

126 YVILRR-----LTKTML-----QOIGRNTYNNPDPIDISHR-----LVI 162

270 QOFLRKORKEAPYDTIQVLDVLRDKPSNDVYVSGRSPFHTSLGKADBGELGDEY 329

163 WFGFTSLIQVENSIMLCTDVSHKVLRSFTVL-DMFNFYHQTEBK-----FQEQVSK 216

330 WAGYFOSLALTMGSLINDVARSFYERIVTDTFSKFLNIRDNLNRPDSRLKVKV 389

217 LIGL-VLTKY-NKTYRVDIDWDONPKSTKADGSEVFLYYRQYNOETDLKOP 274

```

Qy 275 VLVSQPKRRRGGGGLPGRAMLPELC-----VLGTGIDGRNDPNNMKDLAVHTRLT 327
Db 450 AIGT-----GSDTRFVYLPMLCQIDSGQRYTKRINERQ-----VTALLKATCQR 494
Qy 328 PEQOREVRLIDYIKNDNVORELDMGLSPDSNLSFGRILOTE--KHQGGKTFDY 385
Db 495 PDRRENSIKLVYKNNYNDLSKE--FGMSYTTQASIEAVLPPMLKYHDSGKERVY 551
Qy 386 NPQPADMS-----KTRGAPLISVKPLDMLL--YTRRYEANSIQLVFKVTPAMGQ 439
Db 552 NPELGQNMMDIKQNVNGAKVTS-----WTCVSFSTRIDRLPQEFCKQIQGVSKGME 605
Qy 440 MKKA-----IMIEVDRTYAVLRVLOQKVTADTQIVYCLLSNRKDKYDAIKKYCTDCP 494
Db 606 FKQRAIPPTISCPREHIEALDI--HKRAGQLLIYVL-PDYNGSYKIKRIETELG 662
Qy 495 TBSQCVAVARTLQKQTVMAIKTALQNNCKMG-----ELMRVDIPL--KLWIVGI 545
Db 663 IVSQCCQPRQVKNLN--KQYMERVALKIKVYKGRVTVLNDAIRIRIPLITDRPTIMGA 720
Qy 546 DCHMDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFODRGQELVDGLKYLQALR----- 598
Db 721 DVTHPQPEDSSEPSIAAVVASHMDEPEIKIRGLVSAQAREELIQDLVYLVDPQKGLVH 780
Qy 599 -----AMNSCNEWPSRIIYVRDVGQDLKTLVNEVVPQFLDCKSIGRGYNRL 649
Db 781 SGLIREHFLAFRATQIPIQRIIFYRDGVEGQFQVLHEMTAIRKACNSIQENYVPRV 840
Qy 650 TVTVYKKNRNTREFPAGSG-----RLQPLRGTVIDVETVRPEWDFEIVSQAVRSGSV 703
Db 841 TFIYVQKRHTLFPQCHARDMTDSGNIQPTVVDYIICHNEPFLNSAIGQGIS 900
Qy 704 SPHYNVYDNGSLKPDHIOQLTYKLCHIIYNNPVGIVRVPACQYAHKLA 754
Db 901 RPAHYVLVDENGFADQQLMNLNLCYVACTSVSVLPAPAYALVAF 951

```

## RESULT 8

```

T52134
Zwille protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52134
R:Laux, T.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z25975
A:Accession: T52134
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-988 <LNU>
A:Cross-references: EMBL:AJ223508; PIDN:CAA11429.1
C:Genetics:
A:Gene: zwille
A:Map position: 5
C:Superfamily: rabbit translation initiation factor eIF-2C

```

Query Match 12.0%; Score 492; DB 2; Length 988;

Best Local Similarity 23.6%; Pred. No. 2.3e-27;

Matches 204; Conservative 133; Mismatches 397; Indels 132; Gaps 32;

5 VNTRONLDHYKESKSSGIIIVLSTNHFLTSRPMALQYHIDVNEPMEA-----RLRSALLFQHEP-LIGK 63

120 MGRKNSNFAPRGFGTLGKVCYKANHF-LADLPFKDLNQDVYITPEVSSKSVRAII 178

64 -----LFQHEDLIGKCHAPD-----GTLFLPKRLQOKVTE-----VSKTRNGEDVY 106

179 AEIVRLYKESDLRRRLPAYGRKSLYAGLPLFTWKEFSYKIVDEDGILNGFKERASYK 238

107 ITTT-----LTNELPPTSPCLQFNIIFRLIKIMNLQOIGRNTYNNPDP 153

239 VALKTVARAMHHHIGETLAKRADCGQENQIIDIVLR--LSYKRPCVPYGRSFFSPD--I 295

QY 154 DIP---SHRLVWPGFTTSLIQYENSIMLCTDV--SHKVLRSFTYLDMPNFYHTEBHK- 208  
 DB 296 KTFORLGELESWCGFQSIIRPTQMGSLNIDMWSAAFIPELPIEFVLAQLLGKDVLSKP 355  
 QY 209 ---FOEQVSKELIGLVLTAKN---NKTYNVDIDMDQNKSTFKKADSEV--SPLEY 260  
 DB 356 LSDSRVAVIKKGLGVNKEVTHANVRKXVAGLITQPTRELMFPPDENOTKMSVLEYF 415  
 QY 261 RKQINQETIDLKQVYL--VQPRRRRGPGTLPGRAMLPE--LCYLTGLTDMKRDVPMK 318  
 DB 416 QEMVFTTQHTHLPCLQVGNQK---ASYLPMKCKVEGQRYTKELNEQ--- 463  
 QY 319 DLAVITRLTPEORREVRLLIDYI---HKQDNVQRELDWGLSPDSNLSFGRILOTE 374  
 DB 464 ---ITALKTCQRAEGQNDILRTVQNNAYDQDPYAKERGNNISEKLSVEARLLPAP 519  
 QY 375 ---KIHQGGKTFDYNOFADWSKETRGADLISVKPLDNLIIYTRNTE--AANSLIONLF 430  
 DB 520 WLYKHENKCKDCLPQVQGMNMMNK--KXINGMTVSRWACVNFERSVQENVARGFCNELG 577  
 QY 431 KTFPAMGQMRKAIMIEV---DDETE---AYLRVLQCKVADTQIVVCLISNKKDK 481  
 DB 578 QCEVSGEMENPEPVIPIYSARPQVVEKALNHYHTSNKKKKELELLAILPDNNGSL 637  
 QY 482 YDAIKKYLCTDCPTSPQCVAR---TLGQQTVAALATKIALQNNCKRGE---L 530  
 DB 638 YGDLKRICETELGLISQCLTKHVPKISKQ---YLADVSLKINVMKGRNVTLYDAIS 692  
 QY 531 WRV---DPLKLVITVIGDCHMTA--GRSITAGFPASIN--BGMTRMFCRCIFDRCQ 583  
 DB 693 CRIPVSDIP--TIFGADVTHPENGESSPSSIAAVALSODMEVTKIAGLVCAQAHKQ 749  
 QY 584 ELVLDLKYCLQALALR-----AANSCHEWMSRIIVYRDVGQDLKTLVNYE 630  
 DB 750 ELIQDLYKTMQDPVNGVTSQGMIRDLISFRKAGQKPLRIIFRDVSGEQFYQVLYTE 809  
 QY 631 VPQFLDCKSIGRGYNPLVIIVKRVNTRFPA-----OSGRLQNPFGTYIDVET 684  
 DB 810 LDAIRKACASLEPNQOPVTFIVQKRHTRLFPANNHRDNKSTRSGNIIPLGYVADTKIC 869  
 QY 665 RPEWDEPIVGAARVSGVSPTHVYVINDSGLKPDHIGLTYKLCIHYWMPGVIRPA 744  
 DB 870 HPTEDFYLCSHAGIGTISRPAHIVLMDENNFRADGIGSLTNMLCYTAACRVSVEIVP 929  
 QY 745 PCQYAHKLAFLVGSIHREPNLSLN 770  
 DB 930 PAVYAHLAAPRA--RFYLEPEIMQDN 953

Matches 185; Conservative 170; Mismatches 340; Indels 173; Gaps 34;  
 QY 20 GSGGIVLSTVHFL--TSRPMALYQYH-----DYNPLWEARLRALLFQ-----HE 68  
 DB 63 GTRGQKIPLLTHFHKVDVANLQGHFHHYSVALFYDDGP--VEQKVGKILDKHQTYS 121  
 QY 69 DLIGKHAPDG-----TILFPRQLQ--QKTEVFSKTR-----NG----- 102  
 DB 122 DLDGEFAIDGSKITLFTYALPDSNMDSVYLEVSATSKQFVSRRANGSPNGESPDS 181  
 QY 103 -----EDVRIITLTNEIP-----PSPFLOQYNTIFRLLKIM 137  
 DB 182 GDRKLRRENRKSNRVEISYAKIPLQALANAMRGSENSQEARIVLDIILRQHAARQ 241  
 QY 138 NLQIGRNYNPNNDPIDP--SHRLVWPGFTTSLIQYENSIMLCTDV--SHKVLRSFTVL 194  
 DB 242 GCLVRAQSPFH--NDPNCBPVGNLIGCRGHSSFRITQGGSLMDVTTMIIKFPVY 300  
 QY 195 DMENFPTHTEBHKQ--EQVSKELIGLVLTAKNKTAVDDIDMDQNKSTFK----- 247  
 DB 301 DFLANQONARDPYSIDWSKAKRTLNLRVKSPSQGEFTIGLSDKPREQTEELKKNP 360  
 QY 248 ---KADGSEVPLEYRK-----QYNOETDLKQVYLVSQPKRRRGPGTLPGRAML 297  
 DB 361 NENGEFETTEVADVDFRTHRIDQYSADLCIN---VGRPKR-----PTIIP 406  
 QY 298 PELC-----YLTGLTDRNDFNVMKDIAVHRLTPEORREVRGLIDYTHQDNVOR 350  
 DB 407 LELCALVPLQRYTKALTFFQRS-----ALVEKRSQKQERMTVLSKLV--SNYDAP 458  
 QY 351 ELRDMGLSPDNLSSGRILOTETKHQ--GGKTEDVNOQFADWSKETRGADLISVPEPD 408  
 DB 459 LRSQGISISSNFTQVEGRVLPAPLKNGCGSEITPRKGRMPPNKE-----FVEPTKQ 513  
 QY 409 NMLIYTRANYEA---ANSLIONLFKVPAMQMRKAIMIEVD-----RTEAY 455  
 DB 514 RNVVV---NFSARCNVRQVVDLTKIGSKGIEIASPQVFEENQOFRRAPMIRVENM 569  
 QY 456 LRYLQCKVADTQIVVCLISNR--KDKYPAIKKYLCTDCPTSPQCVVATLKKQQTVAI 514  
 DB 570 FKDIQGLKPGVQFICVLPDKKNSDLYGPMKKULTTEGIYVQCAFPQPNQDL--- 626  
 QY 515 ATKIALQNNCKKGG--ELMRVD-----IPLKLVITVIGDCHMTAGRRSIAGV 562  
 DB 627 -TNLLKINAKGGLNSMLSVETRAFTVYSNP---TILLMDVSHG--SPQSDVPSIA 681  
 QY 563 ASINEG---MTRMFSRCIFDROGEIYDGL-----KYCLQALPAMNSCHEWMS 609  
 DB 682 AVVSRREWPLISKYRASVTRQPSKAMETSLVKKNGTEDDGIIKELVDFYSSNRRKE 741  
 QY 610 RIIVYRDVGQDLKTLVYEVPOFLDCKSIGRGYNPLVIIVKRVNTRFPAQSGGR 669  
 DB 742 HILFRDVGSESPNOVLNIEIDQIIEACKLIDANPNRFLLVAKNNHTTFQPTSP- 800  
 QY 670 LQNPFGTYIDVETPEWYDPEFTVSQAVRSGVSPTHVYVINDSGLKPDHIGLTYL 729  
 DB 801 -ENVPQGITIIDNKKIHPKNDPFLCAHAGMIGTRPTHVHYLDEIGFSADELQELVHSL 859  
 QY 730 CHLYWMPGVIRPAPCOYAHKLAFLV 757  
 DB 860 SYVQGSTSATSVAVICTAHLAQAQ 887

RESULT 9  
 A:Accession: A84668  
 A:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: A84668  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84620; MUID:20083487; PMID:10617197  
 A:Accession: A84668  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-930 <STO-  
 A:Cross-References: GB:AE002093; NID:9388534; PIDN:AACT7862.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g27040  
 A:Map position: 2  
 C:Superfamily: rabbit translation initiation factor eIF-2C  
 Query Match 11.9%; Score 488.5; DB 2; Length 930;  
 Best Local Similarity 21.3%; Pred. No. 3.8e-27;

RESULT 10  
 J:Accession: J06569  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 21-Jul-2000  
 C:Accession: J06569  
 R:Zou, C.; Zhang, Z.; Wu, S.; Osterman, J.C.  
 Gene 211, 187-194, 1998  
 A:Title: Molecular cloning and characterization of a rabbit eIF2C protein.  
 A:Reference number: J06569; MUID:98267198; PMID:9602122

A:Accession: PC6505  
 A:Molecule type: protein  
 A:Residues: 336-359;380-694-711 <ZOU1>  
 A:Experimental source: liver  
 A:Accession: JG6569  
 A:Molecule type: mRNA  
 A:Residues: 1-813 <ZOU2>  
 A:Cross-references: GB:AF005355; NID:g253158; PIDN:AC2423.1; PID:g3253159  
 C:Superfamily: rabbit translation initiation factor eIF-2C  
 C:Keywords: liver

Query Match 11.8%; Score 482; DB 2; Length 813;  
 Best Local Similarity 23.9%; Pred. No. 9,1e-27;  
 Matches 191; Conservative 147; Mismatches 341; Indels 120; Gaps 32;

39 PGMALYOHYDNPMEARRLSAL--LFGH--EDLIG-KCHAFDG-----TILFLPKR 87  
 4 PKIDYHIELDKPEKCRVRVREVEHMGHFKAQIGDKRPVDSGRKNLYTMPLPIG 63  
 88 LQOKTEVFSKTRNGED--VRIIT-----TITNELPPTSCQFTNIIFRR 132  
 64 REKVELEVTLEGEKDKRIFKYSIKWBSVSLQALHDLISGRLEVPETIOLADVWRH 122  
 123 LKINNLQOIGRNNYNNPNDIPIDP-SHRLVIMPFTTSILOEENSIMLCTDVSFKVL-RS 190  
 123 -LPSNRVYTPVGSRFTASGCSNPLGGREVFHFGQVRSLSMKMMLNIDVSATAPYKA 181  
 191 ETVLDF--MNFYHQTENK-----FOEVSKEELIGLVITKY--NNKTYRVDIDMD 239  
 192 QVIVFVECEVLDLFSIEGQKPLTDSQKVFKEIKGLKVEITHGQMKRKRVCNVR 241  
 240 QNPKSTFPKADS---EVSFLEYRKYQNGEITLQKPVV-VSPKRRRRGGTLPEPA 294  
 242 PASHOTFPLQSGSGQVTECTVAQYFRDKRKLVRPHPLCQVGEQGH-----T 291  
 295 MLIPFLCYLTG-----LTDKRNDFVWKDLAVHTLTPGEORREYGLIDYHKN 347  
 293 YLPFLVCNIVAGQRCIKULTD---NQTSM--IRATASAPD-ROEELSKLMR--SASFN 343  
 348 VORELDWGLSPDSNLSFSGRILOTETKHOGGKTFDYNPOPADMSKETRAPLISVPL 407  
 344 TDPYRBERFGIMVKDENTDVTGRVLQPPSILYGR-----NKALATPVGVWDM 391  
 408 DN-----WLLIYTRANYEA---NSLIONLFFVTPPAMGMOKR-----AIMIEV 448  
 392 RKKOHTGIEIKVMALACAPORCTEVLKSFTEQLKISPDAGMPICQGPCFCNVAQG 451  
 449 DRTAYLRVLTQOKXTADTQIVVCLLSNRKDKYDAIKKYLCTDCTPSCCVVATLQK 508  
 452 ADSVGRMFRHL-KNTYAGLQLVVIL--PKTPTVYAEVKEVGVDTVLGMAATOCOMKNV--Q 507  
 509 QTVMAIATKIALQNMCKKGG-----ELMRVDIPLKLVIVISIDCYHMTG--RASA 559  
 508 RTTPOTLSNLCKINVKGGVNNIILPGCRPPVPOQVIFLZADVTHP-PADEGKKPSIA 566  
 560 GTVASINSGMTWFRFCIFQDRGOELVLDGLKVCLOALRAMNSCNEVPSRIIYVRDVG 619  
 567 AVVGSMDAPNRYCATVRVQQRHQLIDOLAMVRELLIQFYKSTRFPTRIILFYRDVS 626  
 620 DQGLKLVVYEVPOFLDCKSIGRGVNPILYIVYKKNVRFPA-----QSGRLQNLPL 675  
 627 EOPQOQVHHEHLLAIERACIKLEKYQDPITFIWQKRHTRLFTDKNERYVSGKNLPA 686  
 676 GTVIDEATREPMYDFEIVSOAVRSGSVPTHYNYLYNSGLKPHIQLRTYKLCHIIYN 735  
 687 GTTVDTKIHPTREFDYLCSHAGIOGTSRPSHYHLMMDNRRSDELQILFYQLCHTYVR 746  
 736 MEGVTRVVPAPQYAKLAF 754  
 747 CTRSVSIPAPAYAHVAF 765

RESULT 11

T01113  
 translation initiation factor eIF-2C homolog T2114.12 - Arabidopsis thaliana  
 N:Alternate names: Argonaute (AGO1)-like protein  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Feb-2001  
 C:Accession: T01113; D84739  
 R:Rounsfeld, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
 submitted to the EMBL Data Library, December 1997  
 A:Description: Arabidopsis thaliana chromosome II BAC T2114 genomic sequence.  
 A:Reference number: Z14209  
 A:Accession: T01113  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-887 <ZOU>  
 A:Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702284  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsfeld, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; WUID:20083487; PMID:1061197  
 A:Accession: D84739  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-887 <STO>  
 A:Cross-references: GB:AE002093; NID:g2702284; PIDN:AA91987.1; GSPDB:GN00139  
 A:Genetics:  
 A:Gene: At2g32940; T2114.12  
 A:Map position: 2  
 A:Interons: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3; ;  
 C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.6%; Score 434; DB 2; Length 887;  
 Best Local Similarity 22.2%; Pred. No. 3.2e-23;  
 Matches 198; Conservative 153; Mismatches 349; Indels 190; Gaps 41;

3 FGVTNRQNLDHVESKTSKSGIIVRLSTNHFRLTSR-POMALYOHYH----DYNPLMEAR 57  
 24 YDITTRGV-----GTTGNPIELCTNHNFNVSVPQDVVFQYVYITTEGDAVDGT 75  
 58 RLRSALIFQ-----HEDLICKAFDG-TLFLPKLQK-----YREVSXTANG-- 102  
 76 GISRKMDQLEFKTYSDDLQKRLAYDGEKTLVYVGPLPQNEFPLVIVSGSFSKDCGS 135  
 103 -----EDVRITITLNLPL-----PISPTCLQFYNI 129  
 136 DGGSSGTCRSKRSRFLPRRYKQIHAAEIPKTVLGTQRGAVTDPKSAQDALRVLDIV 195  
 130 FRRLIKIMNLQIGRN-----YNNPNDIPIDISHRLVYMPGTTISLOEENSIMC 180  
 196 LR-----QQAERGCILVRQAFPHSDGHPKVGAGVIGI-RGLHSFRPTGGSLN 246  
 181 TDVS-HKVLSEFVLDPMFNFYHQTENKQ-----EVSKEELIGLVITLRYNNKYRVVD 235  
 247 IDVSTWILTEFGEVIEFL-KAQSVETPRQIMIKVAAKMLGMMVKATHRMMEKRI-- 302  
 236 IDWDONPKS-----TFKKADS-----EVSFLEYRKYQNGEITLQKPVV-VSPKRRR 284  
 303 IGLSKRCNQQLPSMKIKDGEREVPRIETIVDYFKQYIEPISSAVFPLDVGKDR-- 360  
 285 GPGGTLGPMMLPELCYLTGLDVKRNDENWKDLAVNHR--LTGEORREYGLIDYHKN 337  
 361 -----FVYLPLEFENLVSLQ-----RYTKPLSGRQVLLVSSRKAPLERIKITLN 405  
 338 --LIDYHKNQVQRELDWGLSPDSNLSFSGRILOTETKHOGGKTFDYNPOPADMSKE 395  
 406 DAHTYCYDQDPF--LAGGGISIEKEMTQVEGRVLPML-KFGKNEEDFQPCNGRNFEN 461  
 396 TRGAPLISVPLDNWMLIYTRYAYE--ANSLIONLFFVTPPAMGMOM-RKALMIEVD-- 449  
 462 NK-MLEPRAIKSMALV---NFSPPCDSSHISRELISCGMRKGLFIDRPPALVEEDPQ 515



A:Cross-references: EMBL:Z68300; PTDN:CAA92619.1; GSPDB:GNO0022; CESP:T22B3.2b  
 A:Experimental source: clone T22B3  
 C:Genetics:  
 A:Gene: CESP:T22B3.2b  
 A:Map position: 4  
 A:Insertions: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 796/2; 964/2; 1006/1  
 C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.5%; Score 429.5; DB 2; Length 1035;

Best Local Similarity 22.1%; Pred. No. 8.6e-23;

Matches 203; Conservative 144; Mismatches 367; Indels 203; Gaps 36;

QY 20 GSSGIIVLSTNHFRL-TSRPQWALYOHID-YNPLMEARL-RSALIF-----QHED 69  
 DB 91 GTTGKRIIPVKSNEFADLNKPKVAVVQYHVEIHP--GCRKLDKDEKRIIFKAVSDHPN 148  
 QY 70 LIKG--CHAFDGT-ILFLPKRLQ-----QKTEVFSKTRNGEDVRIIT----- 109  
 DB 149 IFHNKFAIADGAHQLYTVAARLEFPDQGSVRLDCEATLPKQNRDRTRCAISQNVGVL 208  
 QY 110 -----TLTNELPPTSPCTLOFYNIIFFRLIKINLQOIGRNYN-----PNDPI 153  
 DB 209 LEMQRTTNLDERVLPTQIIDIICROSLTCPLKN-SANFTYKSSCTRIPTAAGAL 267  
 QY 154 DIPSHRLVWPGETTSILOYEN-SIMLCTDVSHTKLRS--TVLDPMFNHYQ----- 203  
 DB 268 DLEGGK-EMWTGFFSSAHIASYRPLNIDVAHTAFYKTRITVLOFMCQDLNERTSKPNR 326  
 QY 204 -----TEEH 207  
 DB 327 NNPRPGAPGGYRGGRGAGGSGYONFGNRGPGANVDDPGNGLFTMDTSLSDTOLS 386  
 QY 208 KROEOVSKELI-GLVVLTKYNNKTYRVDIDMDQNP--KSTFKKAD--GSEV--SFL 260  
 DB 387 SEFTIIFDSDISGMKIRATHRNPAIRVYKNSQLPADKLMFGQIDEGROVYCSVADYF 446  
 QY 261 RKQYNOEITDLKQVLYSQPKRRRPGGTLPGRAMLIP-ELC-----YLTGLTDKRN 312  
 DB 447 SEKYG--PLKYPKL--PCLHVP--PTRNIFLPMHECLIDSPQYKNNKMEKOTS 495  
 QY 313 DENWMDLAVHTRLPQROREVGRLIDYHKNDNVQRELADWGLSPDSNLSFSGRLQ 372  
 DB 496 --AIIKAAVDA--TQREDRIKOLA--AQSFGTDPFLKEFGVAVSSQWETSARVIO 547  
 QY 373 TEKIHGGKTFDYN---PQFADMSKETRGAPLISVPELDMWLLIYTRRNEAANSLION 428  
 DB 548 PPPIMGSGNNRSINPVFPKDSMSMDHOTLYMPATCRSISMIALVDPDQTSIQTFQS 607  
 QY 429 LEKVTPEMGQMRK-----AIMEVDRTAEVLRVLOQKVTADT-QIYVC 472  
 DB 608 LTMKATAMGNPNPRMWDLVKYGSKEDVCTLFTEIAD-----EYRVNTVDCDCTIV 658  
 QY 473 LLSNRKDKYDAIKKYLCTDPTSPQCVARTLKQOYMAIATKIALQONCKMG-- 528  
 DB 659 VLOSKNSDIYMTVKESQDIYHGMSQVLMKNVSR--PTPACANIIILKNNKMGINSR 716  
 QY 529 -----ELMRVDIPLKLVMIYVIGIDCYDMTAGRR--SIAGFVASINEGTRFSPSCI 577  
 DB 717 IYADQITNKYLVDP--TWVVGIDVTHPTQAEKRMAMPVSAIYANVDLLPQSGYANVK 773  
 QY 578 PQDQGEIVDGLKVCQALMAANSQENMPRSIIYRDVGSGQKLTWNEVPOQLDC 637  
 DB 774 VQKCKESVYVTLDAIRERITITRYHTKQKPAHIYVRDVSSEGSSEVLRBEISQIRAK 833  
 QY 638 LKISIGGYNRLTVIVKQKAVNTRFPAQ--SGGLQNPPLPCTVIDEVTRPEVDFPT 693  
 DB 834 CLAIADPRPPIYIVVQKSHARIFCKFENDWVGAKNVPPTVDTDITVBEDEDFVL 893  
 QY 694 VSGQVAGSGVSPTHNVIYDNGSLKPDHQRLLYKCHITYMPPVGIYRPAPOQYAHKA 753  
 DB 894 CSHYGVQGSRRPARHYLLDECKFTADEIONITYGCHYGRCTRSVSIPTPYTADLYA 953  
 QY 754 FLVGSQIHPREPNLSLN 770

DB 954 TRACHIKKGLADNN 970

## RESULT 14

S41013  
 hypothetical protein ZK757.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 18-Feb-2000

C:Accession: S41013

R:Thomas, K.  
 submitted to the EMBL Data Library, December 1993

A:Reference number: S41011

A:Accession: S41013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-958 <THO>

A:Cross-references: EMBL:229121

C:Genetics:

A:Insertions: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 887/2; 929/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; Length 958;

Best Local Similarity 21.6%; Pred. No. 3.2e-22;

Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;

QY 20 GSSGIIVLSTNHFRL-TSRPQWALYOHID-YNPLMEARL-RSALIF-----QHED 69  
 DB 11 GTTGRIIPVKSNEFADLNKPKVAVVQYHVEIHP--GCRKLDKDEKRIIFKAVSDHPN 68  
 QY 70 LIKG--CHAFDGT-ILFLPKRLQ-----QKTEVFSKTRNGEDVRIIT----- 109  
 DB 69 IFHNKFAIADGAHQLYTVAARLEFPDQGSVRLDCEATLPKQNRDRTRCAISQNVGVL 128  
 QY 110 -----TLTNELPPTSPCTLOFYNIIFFRLIKINLQOIGRNYN-----PNDPI 153  
 DB 129 LEMQRTTNLDERVLPTQIIDIICROSLTCPLKN-SANFTYKSSCTRIPTAAGAL 187  
 QY 154 DIPSHRLVWPGETTSILOYEN-SIMLCTDVSHTKLRS--TVLDPMFNHYQ----- 203  
 DB 188 DLEGGK-EMWTGFFSSAHIASYRPLNIDVAHTAFYKTRITVLOFMCQDLNERTSKPNR 246  
 QY 204 -----TEEH 207  
 DB 247 NNPRPGAPGGYRGGRGAGGSGYONFGNRGPGANVDDPGNGLFTMDTSLSDTOLS 306  
 QY 205 KROEOVSKELI-GLVVLTKYNNKTYRVDIDMDQNP--KSTFKKAD--GSEV--SFL 257  
 DB 307 SEFTIIFDSDISGMKIRATHRNPAIRVYKNSQLPADKLMFGQIDEGROVYCSVADYF 366  
 QY 258 RKQYNOEITDLKQVLYSQPKRRRPGGTLPGRAMLIP-ELC-----YLTGLTDKRN 312  
 DB 367 SEKYG--PLKYPKL--PCLHVP--PTRNIFLPMHECLIDSPQYKNNKMEKOTS 495  
 QY 313 DENWMDLAVHTRLPQROREVGRLIDYHKNDNVQRELADWGLSPDSNLSFSGRLQ 372  
 DB 496 --AIIKAAVDA--TQREDRIKOLA--AQSFGTDPFLKEFGVAVSSQWETSARVIO 547  
 QY 373 TEKIHGGKTFDYN---PQFADMSKETRGAPLISVPELDMWLLIYTRRNEAANSLION 428  
 DB 548 PPPIMGSGNNRSINPVFPKDSMSMDHOTLYMPATCRSISMIALVDPDQTSIQTFQS 607  
 QY 429 LEKVTPEMGQMRK-----AIMEVDRTAEVLRVLOQKVTADT-QIYVC 472  
 DB 608 LTMKATAMGNPNPRMWDLVKYGSKEDVCTLFTEIAD-----EYRVNTVDCDCTIV 658  
 QY 473 LLSNRKDKYDAIKKYLCTDPTSPQCVARTLKQOYMAIATKIALQONCKMG-- 528  
 DB 659 VLOSKNSDIYMTVKESQDIYHGMSQVLMKNVSR--PTPACANIIILKNNKMGINSR 716  
 QY 529 -----ELMRVDIPLKLVMIYVIGIDCYDMTAGRR--SIAGFVASINEGTRFSPSCI 577  
 DB 717 IYADQITNKYLVDP--TWVVGIDVTHPTQAEKRMAMPVSAIYANVDLLPQSGYANVK 773  
 QY 578 PQDQGEIVDGLKVCQALMAANSQENMPRSIIYRDVGSGQKLTWNEVPOQLDC 637  
 DB 774 VQKCKESVYVTLDAIRERITITRYHTKQKPAHIYVRDVSSEGSSEVLRBEISQIRAK 833  
 QY 638 LKISIGGYNRLTVIVKQKAVNTRFPAQ--SGGLQNPPLPCTVIDEVTRPEVDFPT 693  
 DB 834 CLAIADPRPPIYIVVQKSHARIFCKFENDWVGAKNVPPTVDTDITVBEDEDFVL 893  
 QY 694 VSGQVAGSGVSPTHNVIYDNGSLKPDHQRLLYKCHITYMPPVGIYRPAPOQYAHKA 753  
 DB 894 CSHYGVQGSRRPARHYLLDECKFTADEIONITYGCHYGRCTRSVSIPTPYTADLYA 953  
 QY 754 FLVGSQIHPREPNLSLN 770



```

QY 575 RCIFQDROGELVDGLKVCLOALRAMNSCNEWMSRIIVYRDGVDGQLKTLVNYEVPQF 634
DB 694 NVKVKCKRESVYLLDAIRERITTFYRHTKQKPARIIYRDGVSQSPSEVLREIQSI 753
QY 635 LDCLKSIGRGYNPRLTVIVYKRVNTRFFAQ---SGRLQNPFGTVIDVEVTRPEWYD 690
DB 754 RTACLAIAEDFRPPIITIVYQKRNHARIFCKYQNDWVGAKXVPGTVDGIVSEBGF 813
QY 691 PFIVSQAVRSGSVPTHTNIVYDNGSLKPRHIGLTLTKLCHIYNNMGVIRVAPCOYAH 750
DB 814 FYLSHVGQGTSPRKHVHLDECKFTADEIOSITIGMCHYGRCTRSVSIPTVYIAD 873
QY 751 KLAFLVGSIHREPNIISLN 770
DB 874 LVATRAACHVARKLGLADNN 893

RESULT 15
DB8568
protein ZK757.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: DB8568
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see web site genome.wustl.edu/gen/c/elegans/ and www.sanger.ac.uk/projects/c_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: DB8568
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1040 <STO>
A:Cross-references: GB:chr_III; PIDN:CA82941.1; PID:93877004; GSPDB:GN00021; CSDP:ZK757
C:Genetics:
A:Gene: ZK757.3
A:Superfamily: 3
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; Length 1040;
Best Local Similarity 21.6%; Pred. No. 3.6e-22;
Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;

```

```

QY 20 GSSGIIVLSTNTHRL--TSRQMALYOYHID--VNPIMEARRL---RSALF-----QHED 69
DB 93 GTIGQIVKSNFPMADIKNPKCVVIOYVHEIHP--GCRKLDXDEKRITFWKAVSDHPN 150
QY 70 LIGK--CHAFDGT--ILFLPKRLQ-----QKTEVSKTRNGEDVARTI----- 109
DB 151 IFHNKFALAYGAHQLYVARLEFPDQGSVRLDCEASLPRKNDRTCAISIQNGFVL 210
QY 110 -----TLNNEPPTPTCLQFYNIIFRLKIMLQOIGRYNN-----PNDPI 153
DB 211 LEMORTNNLDERVLTPIQIIDICRQSLTCEPLKN--SANFYTWKSSCYRIPTAAGAL 269
QY 154 DIPSHRLVYMPGFTTSILOYEN--SIMLCTDVSHKVLASE--TVLDFMENFYHQ----- 203
DB 270 DIEGK--EMWTGFPSSAHASNYRPLNIDVAHTAFYKTRITVLOFMCDVNERTSKPNR 328
QY 204 -----T 204
DB 329 NNPRGPGGPGGPGYRGGRGGGSGYGNFNGRPGANVRDFFGNGLFTMTDLSDRT 388
QY 205 EHNKFOEQYSKELI--GLVVLTKYNNKTYRVDIDMDONP--KSTFKAD--GSEY--SFL 257
DB 389 QUSSEETRIFGDAIKMKTRAHARNRAIRYKYNLSQLPADKLMFGQIDBGRQYVCSVA 448
QY 258 EYRKQYNOEITDLKQPVYVSOQKRRRGGTLPGPAMLIP--ELC-----YLTGLTDK 309
DB 449 DYFSEKYG---PLKYFKL---FCLHVG---PFRNIFLPMHEHCLIDSPQYNNKMSK 497
QY 310 MENDFVNMDLAVHTRLTPEGRQREVRGLIDYIHKNQVQRELARDWGLSPDSNLLSFGSR 369

```

```

DB 498 QTS--AIKAAVDA-----TQREIRIKOLA--AQASFQIDPFLKFGVAVSSOMQTPAR 549
QY 370 ILQTEKINGGKTEPYN-----POFADMEKETRGAPLISVKPLDMLIYTRRYEAANSI 425
DB 550 VIQPPPIFGGNKRVNVVFPKQGSWTMDQYLMPEATCRSYSMLALVDRDQTSLOTF 609
QY 426 IONLEKVTYPMGMQWEX-----AIMEVDRTEAVLRVLOQKRTADT-QI 469
DB 610 CQSLMTKATAGMNPFPWEDLVKYGRSKEDVCTLFETIAD-----EYRVNTEVDC 660
QY 470 VYCLSSNRKQDYDAIKKYKLTDCETPSQCVVARTIGRQYVMAIATKIALQNMCKXGG- 528
DB 661 IIVVLQSKNSDIYMTVKEQSDIVAGIMQCVLAKRVSR--PFPATCAIIVLKLNNKXGI 718
QY 529 -----ELMRVDIPLKLMVIYVIGIDCYHDMTAGR-----SIAGFVASINEGTRNFS 574
DB 719 NSRIVADKTIYKXLYVDP---TMVVGIDVTHPTQAEKMNPSVAIAVANDLLPQSYGA 775
QY 575 RCIFQDROGELVDGLKVCLOALRAMNSCNEWMSRIIVYRDGVDGQLKTLVNYEVPQF 634
DB 776 NVKVKCKRESVYLLDAIRERITTFYRHTKQKPARIIYRDGVSQSPSEVLREIQSI 835
QY 635 LDCLKSIGRGYNPRLTVIVYKRVNTRFFAQ---SGRLQNPFGTVIDVEVTRPEWYD 690
DB 836 RTACLAIAEDFRPPIITIVYQKRNHARIFCKYQNDWVGAKXVPGTVDGIVSEBGF 895
QY 691 PFIVSQAVRSGSVPTHTNIVYDNGSLKPRHIGLTLTKLCHIYNNMGVIRVAPCOYAH 750
DB 896 FYLSHVGQGTSPRKHVHLDECKFTADEIOSITIGMCHYGRCTRSVSIPTVYIAD 955
QY 751 KLAFLVGSIHREPNIISLN 770
DB 956 LVATRAACHVARKLGLADNN 975

```

Search completed: May 5, 2004, 15:14:48  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:10:00 ; Search time 52 Seconds  
(without alignments)  
4702.435 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102  
Sequence: 1 MIFGVNTRQNDHVKESKTG.....VGQSIHREPNTLSNRLYYL 775

Scoring table: BLOSUM62

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  sp. arctear: *
2:  sp. bactera: *
3:  sp. fungi: *
4:  sp. human: *
5:  sp. invertibrate: *
6:  sp. mammal: *
7:  sp. mhc: *
8:  sp. organelle: *
9:  sp. phage: *
10: sp. plant: *
11: sp. rodent: *
12: sp. virus: *
13: sp. vertebrate: *
14: sp. unlabeled: *
15: sp. tvirus: *
16: sp. bacteriap: *
17: sp. arctear: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4102	100.0	775	4	Q96JD5	Q96JD5 homo sapien
2	4087	99.6	861	4	Q96J94	Q96J94 homo sapien
3	4072	99.3	861	4	Q8THY5	Q8THY5 homo sapien
4	4071	99.2	861	4	Q95404	Q95404 homo sapien
5	3980	97.0	882	11	Q9JMB7	Q9JMB7 mus musculus
6	3883	94.7	829	4	Q8NAB0	Q8NAB0 homo sapien
7	2828.5	69.0	858	13	Q8NVX0	Q8NVX0 Brachydanio
8	2144	52.3	882	4	Q723Z2	Q723Z2 homo sapien
9	2103.5	51.3	882	4	Q8NHM2	Q8NHM2 homo sapien
10	2101.5	51.2	882	4	Q8NV88	Q8NV88 homo sapien
11	2098.5	49.0	852	4	Q723Z4	Q723Z4 homo sapien
12	2009.5	45.0	854	5	Q9GPA8	Q9GPA8 strongyloce
13	1859.5	45.3	878	11	Q8GCT6	Q8GCT6 mus musculus
14	1838.5	44.8	866	4	Q8NBG9	Q8NBG9 homo sapien
15	1811.5	44.2	808	5	Q9GPA7	Q9GPA7 strongyloce
16	1766	43.1	971	11	Q9JMB6	Q9JMB6 mus musculus

17	1764	42.8	971	11	08CDG1
18	1754	42.8	973	4	08TC59
19	1468	35.8	866	5	076922
20	1389.5	33.9	580	11	093MW6
21	1295.5	31.6	824	5	P90786
22	1230	30.0	530	4	096SM6
23	1209	29.5	421	5	09GPM6
24	1162	28.3	722	5	017567
25	1059	25.8	978	5	08MNL6
26	1009	24.6	1208	5	086HF5
27	989	24.1	791	5	08IS68
28	981	23.9	791	5	086CV7
29	962	23.5	371	4	09NMZ8
30	958.5	23.4	371	13	07JZM5
31	804.5	19.6	780	5	08MKZ9
32	803.5	19.6	780	5	08MOL1
33	728	17.7	781	5	09U5C9
34	537.5	13.0	1058	10	085LR2
35	520.5	12.7	1194	10	09SMR2
36	508.5	12.4	860	11	08CIG0
37	508.5	12.4	1014	10	09SMF3
38	507	12.4	978	10	08LPT0
39	505.5	12.3	891	5	086B39
40	505.5	12.3	910	5	016720
41	504.5	12.3	861	11	08CJF8
42	503	12.3	896	10	084Y14
43	502.5	12.3	951	11	08BFF4
44	502	12.2	860	11	08CJF9
45	501.5	12.2	1055	10	07Y001
					08cdq1 mus musculu
					08tc59 homo sapien
					076922 drosophila
					093mw6 mus musculu
					P90786 caenorhabd
					096sm6 homo sapien
					09gpm6 strongyloce
					017567 caenorhabd
					08ml6 dictyosteli
					08nfs dictyosteli
					081998 stylyonchia
					086cv7 stylyonchia
					09nmz8 homo sapien
					07jzm5 brachydantio
					08mkz9 tetrahymena
					08mol1 tetrahymena
					09u5c9 paramectum
					085lr2 oryza sativ
					086lf2 arabidopsis
					08cig0 mus musculu
					08smf3 arabidopsis
					08lpt0 oryza sativ
					086b39 caenorhabd
					016720 caenorhabd
					08cjf8 mus musculu
					084y14 arabidopsis
					08bff4 mus musculu
					08cjf9 mus musculu
					07y001 oryza sativ

## ALIGNMENTS

	RESULT	1
ID	Q96JDS	PREDIMINARY; PRT; 775 AA.
AC	Q96JDS;	
DT	01-DEC-2001 (TrEMBLrel. 19, Created)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation updates)	
DE	HIMI.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxId=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=testis;	
RA	Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Minhdud N., Weller K.P., Hoffman R.;	
RT	"Human CD34+ Stem Cells Express the him1 Gene, a Human Homolog of the Drosophila Gene plw.";	
RL	Submitted (May-2000) to the EMBL/GeneBank/DDBJ databases.	
DR	EMBL; AF284004; AAA92281.1; -.	
DR	InterPro; IPR003100; PAZ.	
DR	InterPro; IPR003165; Plwl.	
DR	Pfam; PF02170; PAZ; 1.	
DR	Pfam; PF02171; Plwi; 1.	
DR	PROSITE; PSS0821; PAZ; 1.	
DR	PROSITE; PSS0822; PIWI; 1.	
SQ	SEQUENCE 775 AA; 89484 MW; DFI69A2B9EAFD916 CRC64;	
<hr/>		
Query Match	100.0%; Score 4102; DB 4; Length 775;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0,		
OY	1 MIFFGNTRONLDHYKESKTSSGIIIVRSLTHFULTSNPOMALVNYHNDYNPWMSARLR	60
Dd	1 MIFFGNTRONLDHYKESKTSGGIIVRSLTHFRLTSPQVALIYHDINPLWEARRLR	60
OY	61 SALTFGEEDLGKHAFGGTLIPPKRIQQATVEFSKTRNGEDVRITITLNELPTPS	120

```

Db      61  SALLFQHEEDLIGKCHAFDGTLLFLPKRLQOKVTEVFSEKTRNGEDVATITTLTNELPPTSP 120
Qy      121  TCLOFNILIRRLIKINLGOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 180
Db      121  TCLOFNILIRRLIKINLGOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMC 180
Qy      181  TDVSHKVLRSFTVLDPMENFYHQTBEHKFOQVSKELIGLVLTKNKNTYRVDIDMDQ 240
Db      181  TDVSHKVLRSFTVLDPMENFYHQTBEHKFOQVSKELIGLVLTKNKNTYRVDIDMDQ 240
Qy      241  NPKSTFKADGSEVSEFLEYRKQYNQOEITDLKQPVLSQPRRRPGGTLPGPAMLPEL 300
Db      241  NPKSTFKADGSEVSEFLEYRKQYNQOEITDLKQPVLSQPRRRPGGTLPGPAMLPEL 300
Qy      301  CYLTGLTDKMNNDPVMKDLAVHTRLTPROREVEGRLLIDYIHKNDVQRELRDWSLSPD 360
Db      301  CYLTGLTDKMNNDPVMKDLAVHTRLTPROREVEGRLLIDYIHKNDVQRELRDWSLSPD 360
Qy      361  SNLISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 420
Db      361  SNLISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYE 420
Qy      421  AANSLIONLFVYTPAMGQMKAKAIMIEVDRETEAYLAVLQOKVTAQIYVCLLSNRKC 480
Db      421  AANSLIONLFVYTPAMGQMKAKAIMIEVDRETEAYLAVLQOKVTAQIYVCLLSNRKC 480
Qy      481  KYDAIKKYLCTDCTPSCQVAVARTLGKQVMAIATKIALQNCCKMGELMRVDIPLKLY 540
Db      481  KYDAIKKYLCTDCTPSCQVAVARTLGKQVMAIATKIALQNCCKMGELMRVDIPLKLY 540
Qy      541  MIVGIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAY 600
Db      541  MIVGIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAY 600
Qy      601  NSCNEYMSRIIYVRDVGQDLKTLVNYEVPQFLDCLKSIGRGYNPLTVIVKKRVNT 660
Db      601  NSCNEYMSRIIYVRDVGQDLKTLVNYEVPQFLDCLKSIGRGYNPLTVIVKKRVNT 660
Qy      661  RFPAGSGRLQNPDPGTVIDEYTRPEWYDFPIYSQAVRSGSVPTTHNYIYDNGKLKPD 720
Db      661  RFPAGSGRLQNPDPGTVIDEYTRPEWYDFPIYSQAVRSGSVPTTHNYIYDNGKLKPD 720
Qy      721  HIORLTYLCHIIYNNPVGIRVPACQYAHKLAFLVGSIHREPLSLSNRLYYL 775
Db      721  HIORLTYLCHIIYNNPVGIRVPACQYAHKLAFLVGSIHREPLSLSNRLYYL 775

RESULT 2
Q96J94  PRELIMINARY; PRT; 861 AA.
AC      096J94;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE      PIWI protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RA      Sha J.H.;
RT      "Cloning and identification of human piwi protein related to testis
RT      development."
RL      Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AF387507; AAK69348.1;
DR      InterPro; IPR003100; PAZ.
DR      InterPro; IPR003165; PAZ.
DR      Pfam; PF02170; PAZ; 1.
DR      Pfam; PF02171; PAZ; 1.
DR      PROSITE; PS50821; PAZ; 1.
DR      PROSITE; PS50822; PAZ; 1.

```

```

SQ      SEQUENCE      861 AA; 98603 MM; 58D76C7321DEFA4 CRC64;
Query Match      99.6%; Score 4087; DB 4; Length 861;
Best Local Similarity 100.0%; Pred. No. 1e-317;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4  GYNTFRONLDHYKESKTSSSGIIIVRLSTNHFPLTSRPGMALYQYHIDVNPMLRRLRSAL 63
Db      90  GYNTFRONLDHYKESKTSSSGIIIVRLSTNHFPLTSRPGMALYQYHIDVNPMLRRLRSAL 149
Qy      64  LFOHEDLIGKCHAFDGTLLFLPKRLQOKVTEVFSEKTRNGEDVATITTLTNELPPTSPCL 123
Db      150  LFOHEDLIGKCHAFDGTLLFLPKRLQOKVTEVFSEKTRNGEDVATITTLTNELPPTSPCL 209
Qy      124  QFYNIIFRLLIKINLGOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMLCTDV 183
Db      210  QFYNIIFRLLIKINLGOIGNNYNNPNDPIDIPSHRLVIMPGFTTSLQYENSIMLCTDV 269
Qy      184  SHKVLRSFTVLDPMENFYHQTBEHKFOQVSKELIGLVLTKNKNTYRVDIDMDQMPK 243
Db      270  SHKVLRSFTVLDPMENFYHQTBEHKFOQVSKELIGLVLTKNKNTYRVDIDMDQMPK 329
Qy      244  STFKADGSEVSEFLEYRKQYNQOEITDLKQPVLSQPRRRPGGTLPGPAMLPELCYL 303
Db      330  STFKADGSEVSEFLEYRKQYNQOEITDLKQPVLSQPRRRPGGTLPGPAMLPELCYL 389
Qy      304  TGLTDKMNNDPVMKDLAVHTRLTPROREVEGRLLIDYIHKNDVQRELRDWSLSPDNL 363
Db      390  TGLTDKMNNDPVMKDLAVHTRLTPROREVEGRLLIDYIHKNDVQRELRDWSLSPDNL 449
Qy      364  ISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYEAA 423
Db      450  ISFSGRLIQTEKIHQGGKTFDYNQFADWSKETGAPLISYKPLDNLIIYTRRNYEAA 509
Qy      424  SLIONLFVYTPAMGQMKAKAIMIEVDRETEAYLAVLQOKVTAQIYVCLLSNRKDYD 483
Db      510  SLIONLFVYTPAMGQMKAKAIMIEVDRETEAYLAVLQOKVTAQIYVCLLSNRKDYD 569
Qy      484  AIKKYLCTDCTPSCQVAVARTLGKQVMAIATKIALQNCCKMGELMRVDIPLKLVNI 543
Db      570  AIKKYLCTDCTPSCQVAVARTLGKQVMAIATKIALQNCCKMGELMRVDIPLKLVNI 629
Qy      544  GIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAY 603
Db      630  GIDCHDMTAGRRSIAGFVASINEGMTWFSRCLFQDRGQELVDGLKVCLOALIRAY 689
Qy      604  NEYMSRIIYVRDVGQDLKTLVNYEVPQFLDCLKSIGRGYNPLTVIVKKRVNTRFF 663
Db      690  NEYMSRIIYVRDVGQDLKTLVNYEVPQFLDCLKSIGRGYNPLTVIVKKRVNTRFF 749
Qy      664  AOSGRLQNPDPGTVIDEYTRPEWYDFPIYSQAVRSGSVPTTHNYIYDNGKLKPDHQ 723
Db      750  AOSGRLQNPDPGTVIDEYTRPEWYDFPIYSQAVRSGSVPTTHNYIYDNGKLKPDHQ 809
Qy      724  RLTYLCHIIYNNPVGIRVPACQYAHKLAFLVGSIHREPLSLSNRLYYL 775
Db      810  RLTYLCHIIYNNPVGIRVPACQYAHKLAFLVGSIHREPLSLSNRLYYL 861

RESULT 3
Q87B5  PRELIMINARY; PRT; 861 AA.
AC      Q87B5;
DT      01-JUN-2002 (Tremblrel. 21, Created)
DT      01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT      01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE      PIWI-like 1 (Drosophila).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.

```

RC TISSUE=Testis;  
 RA Struhsberg R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC028581; AAH28581.1; -.  
 DR Genew; HGNC:9007; PIWIL1.  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; PIWI.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; PIWI; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; PIWI; 1.  
 DR PROSITE; PS50822; PIWI; 1.  
 SQ SEQUENCE 861 AA; 98545 MW; D33376EDED743A CRC64;

Query Match 99.3%; Score 4072; DB 4; Length 861;  
 Best Local Similarity 99.6%; Pred. No. 1.5e-316;  
 Matches 769; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

4 GVNTRQNDLVKESKTGSSGIIIVRLSTNHFRLTSRQWALYQYHIDYNPIMEARRLRSAL 63  
 90 GVNTRQNDLVKESKTGSSGIIIVRLSTNHFRLTSRQWALYQYHIDYNPIMEARRLRSAL 149  
 64 LFGHEDLIGKCHAFDGTLLFLPKRLOQKTEVFSKTRNGEDVRLITLTLNELPPTSPTCL 123  
 150 LFGHEDLIGKCHAFDGTLLFLPKRLOQKTEVFSKTRNGEDVRLITLTLNELPPTSPTCL 209  
 124 QFYNIIFRLLKIMNLQOIGRNYNPNPDIPISSHRLVYWGFTTSLQYENSIMLCTDV 183  
 210 QFYNIIFRLLKIMNLQOIGRNYNPNPDIPISSHRLVYWGFTTSLQYENSIMLCTDV 269  
 184 SHKVRSETVLDPMNPFYHQTSEHKFOEVSKEILGLVYLTKNNKTYRVDIDMDQNPX 243  
 270 SHKVRSETVLDPMNPFYHQTSEHKFOEVSKEILGLVYLTKNNKTYRVDIDMDQNPX 329  
 244 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLLPGPAMLIPELCYL 303  
 330 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLLPGPAMLIPELCYL 389  
 304 TGLTDKRRNDPMNMDLAHTRLTPEQRORVGRLLIDYIHKNDNVQRELDMGSLSPDNL 363  
 390 TGLTDKRRNDPMNMDLAHTRLTPEQRORVGRLLIDYIHKNDNVQRELDMGSLSPDNL 449  
 364 LPSGRILLOTEKIHOGKTEFDYNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAA 423  
 450 LPSGRILLOTEKIHOGKTEFDYNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAA 509  
 424 SLIQNLFKYTPAMGMQMRKAIMIEVDRTAVALYLOQKTAADTOIYVCLLSNRKDXD 483  
 510 SLIQNLFKYTPAMGMQMRKAIMIEVDRTAVALYLOQKTAADTOIYVCLLSNRKDXD 569  
 484 AIKKYLCIDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGSELNRVDIPLKLVIV 543  
 570 AIKKYLCIDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGSELNRVDIPLKLVIV 629  
 544 GIDCYHDMTAGRRSLAGFVASINEGTRWFSRCIFQDRGQELVLDGLKYLQALALPAMNSC 603  
 630 GIDCYHDMTAGRRSLAGFVASINEGTRWFSRCIFQDRGQELVLDGLKYLQALALPAMNSC 689  
 604 NEYMPRIIIVRDGVDGQKLTLVNEVFPQPLDCLKSIRGNPRLIYIVAKRVNTPF 663  
 690 NEYMPRIIIVRDGVDGQKLTLVNEVFPQPLDCLKSIRGNPRLIYIVAKRVNTPF 749  
 664 AOSGGLONPLPGTVIDEVTRPEWYDFIVSOAVRSGVSPTHNVIVYDNGSLPDIHQ 723  
 750 AOSGGLONPLPGTVIDEVTRPEWYDFIVSOAVRSGVSPTHNVIVYDNGSLPDIHQ 809  
 724 RLTYLKGILYNNPVGIVRPAQCQVAHKLAFIVGOSIHREPNLSISNRLYYL 775  
 810 RLTYLKGILYNNPVGIVRPAQCQVAHKLAFIVGOSIHREPNLSISNRLYYL 861

AC 095404;  
 DT 01-MAY-1999 (TRENBERG). 10, Created)  
 DT 01-MAY-2002 (TRENBERG). 20, Last sequence update)  
 DT 01-JUN-2003 (TRENBERG). 24, Last annotation update)  
 DE HIWI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=99069219; PubMed=9851978;  
 RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;  
 RT "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal.";  
 RL Genes Dev. 12:3715-3727 (1998).  
 RN (2)  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Qiao D., Zeeman A.-M., Deng W., Locifenga L.H.J., Lin H.;  
 RT "Molecular characterization of hiwi, a human member of the piwi stem cell gene family whose overexpression is correlated to seminomas.";  
 RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF104260; AAC97571.2; -.  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; PIWI.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; PIWI; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; PIWI; 1.  
 DR PROSITE; PS50822; PIWI; 1.  
 SQ SEQUENCE 861 AA; 98530 MW; 43D7F60B9D997B7 CRC64;

Query Match 99.2%; Score 4071; DB 4; Length 861;  
 Best Local Similarity 99.7%; Pred. No. 1.8e-316;  
 Matches 770; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

4 GVNTRQNDLVKESKTGSSGIIIVRLSTNHFRLTSRQWALYQYHIDYNPIMEARRLRSAL 63  
 90 GVNTRQNDLVKESKTGSSGIIIVRLSTNHFRLTSRQWALYQYHIDYNPIMEARRLRSAL 149  
 64 LFGHEDLIGKCHAFDGTLLFLPKRLOQKTEVFSKTRNGEDVRLITLTLNELPPTSPTCL 123  
 150 LFGHEDLIGKCHAFDGTLLFLPKRLOQKTEVFSKTRNGEDVRLITLTLNELPPTSPTCL 209  
 124 QFYNIIFRLLKIMNLQOIGRNYNPNPDIPISSHRLVYWGFTTSLQYENSIMLCTDV 183  
 210 QFYNIIFRLLKIMNLQOIGRNYNPNPDIPISSHRLVYWGFTTSLQYENSIMLCTDV 269  
 184 SHKVRSETVLDPMNPFYHQTSEHKFOEVSKEILGLVYLTKNNKTYRVDIDMDQNPX 243  
 270 SHKVRSETVLDPMNPFYHQTSEHKFOEVSKEILGLVYLTKNNKTYRVDIDMDQNPX 329  
 244 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLLPGPAMLIPELCYL 303  
 330 STEFKADGSEVSFLEYRKQYNOEITDLKQPVLSQPKRRRGGTLLPGPAMLIPELCYL 389  
 304 TGLTDKRRNDPMNMDLAHTRLTPEQRORVGRLLIDYIHKNDNVQRELDMGSLSPDNL 363  
 390 TGLTDKRRNDPMNMDLAHTRLTPEQRORVGRLLIDYIHKNDNVQRELDMGSLSPDNL 449  
 364 LPSGRILLOTEKIHOGKTEFDYNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAA 423  
 450 LPSGRILLOTEKIHOGKTEFDYNPOFADMSKETRGAPLISVKELDNWLLIYTRRYEAA 509  
 424 SLIQNLFKYTPAMGMQMRKAIMIEVDRTAVALYLOQKTAADTOIYVCLLSNRKDXD 483  
 510 SLIQNLFKYTPAMGMQMRKAIMIEVDRTAVALYLOQKTAADTOIYVCLLSNRKDXD 569  
 484 AIKKYLCIDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGSELNRVDIPLKLVIV 543  
 570 AIKKYLCIDCPTSPQCVVARTLGKQVTMAIATKIALQNNCKMGSELNRVDIPLKLVIV 629

QY 544 GIDCYHMTAGRRSIAAGFVASINEGTMFSCIFODRGGLVDGLKVCLOAALRAMNSC 603  
 DB 630 GIDCYHMTAGRRSIAAGFVASINEGTMFSCIFODRGGLVDGLKVCLOAALRAMNSC 689  
 QY 604 NEWPSRIIVYRDSVGGGLKTLVNYEVPOLDLCKSIGRGYNPLTVIVKRVNTRFF 663  
 DB 690 NEWPSRIIVYRDSVGGGLKTLVNYEVPOLDLCKSIGRGYNPLTVIVKRVNTRFF 749  
 QY 664 AOSGGRLONPDPGVIDVEYTRPEWYDFETVSOAVRSGSVSPTHYNYIDNSGLKPDHIQ 723  
 DB 750 AOSGGRLONPDPGVIDVEYTRPEWYDFETVSOAVRSGSVSPTHYNYIDNSGLKPDHIQ 809  
 QY 724 RLTYKLCIHYNNPVGIVRVPAPCOYAHKLAFLVQGSIHREPNLSLSNRLYYL 775  
 DB 810 RLTYKLCIHYNNPVGIVRVPAPCOYAHKLAFLVQGSIHREPNLSLSNRLYYL 861

## RESULT 5

QJMB7 PRELIMINARY; PRT; 862 AA.

AC 09JMB7  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE MIWI (P1wi).  
 GN PIWI1 OR MIWI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP Mlyagawa S.K., Kimura T., Nakano T.;  
 RT "Molecular Cloning and characterization of P1wi family genes."  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Deng W., Lin H.;  
 RT "Mwi, a murine homolog of P1wi, encodes a cytoplasmic protein  
 essential for spermatogenesis."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB032604; BAA33705.1; -  
 DR EMBL; AF438405; AAL31014.1; -  
 DR MGD; MGI:1928897; P1wi1.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0003727; F:single-stranded RNA binding; IDA.  
 DR GO; GO:0007283; P:spermatogenesis; IMP.  
 DR InterPro; IPR003100; P1W.  
 DR InterPro; IPR003165; P1wi.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; P1wi; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; P1WI; 1.  
 SQ SEQUENCE 862 AA; 98574 MW; 4558BD13284CC4C CRC64;

Query Match 97.0%; Score 3980; DB 11; Length 862;  
 Best Local Similarity 96.4%; Pred. No. 3.4e-309;  
 Matches 744; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 4 GNTFONLDHYVESSTGSSGIIIVRLSTNHFRITSRPPALQYHIDYVPLMEARRLSAL 63  
 DB 91 GNTFONLDHYVESSTGSSGIIIVRLSTNHFRITSRPPALQYHIDYVPLMEARRLSAL 150  
 DB 151 LFQHEDLIGRCHAFDGTLLFLPKRLQOQVTEVFSKTRNGEDVTRITITLNELPSPICL 210  
 QY 64 LFQHEDLIGRCHAFDGTLLFLPKRLQOQVTEVFSKTRNGEDVTRITITLNELPSPICL 123  
 DB 124 LFQHEDLIGRCHAFDGTLLFLPKRLQOQVTEVFSKTRNGEDVTRITITLNELPSPICL 210  
 QY 124 QFYNIIFRRLKIKMLQOIGRYVYVNPDPIDPSRLVIYVPGFTSIIQYENSIMLCIDV 183  
 DB 211 QFYNIIFRRLKIKMLQOIGRYVYVNPDPIDPSRLVIYVPGFTSIIQYENSIMLCIDV 270  
 QY 184 SHKVLRSFETVLDPMENFYHQTEHKKFOQSVSKELIGLVVLTFRYNNKTYRVDDIDMDONPK 243

DB 271 SHKVLRSFETVLDPMENFYHQTEHKKFOQSVSKELIGLVVLTFRYNNKTYRVDDIDMDONPK 330  
 QY 244 STFKKADSEVSEFLEYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPMLIPELCYL 303  
 DB 331 STFKKADSEVSEFLEYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPMLIPELCYL 390  
 QY 304 TGLDTPKANDPNNWKDLAVHRLTPEQORQPEVGLTIYHKNQVORELSDMGSPDNL 363  
 DB 391 TGLDTPKANDPNNWKDLAVHRLTPEQORQPEVGLTIYHKNQVORELSDMGSPDNL 450  
 QY 364 LSFSGRLIQTEKIHGGKTFPDYNPQFADWSKETGAPLISVYKPLDMLIYTRNRYEAN 423  
 DB 451 LSFSGRLIQSEKIHGGKTFPDYNPQFADWSKETGAPLISVYKPLDMLIYTRNRYEAN 510  
 QY 424 SLIQVLFVTPAMGMQKALIMIEVDRTETAYLRLVLOQKTAADQIYVCLSSNRKQYD 483  
 DB 511 SLIQVLFVTPAMGMQKALIMIEVDRTETAYLRLVLOQKTAADQIYVCLSSNRKQYD 570  
 QY 544 GIDCYHMTAGRRSIAAGFVASINEGTMFSCIFODRGGLVDGLKVCLOAALRAMNSC 603  
 DB 571 AIKKYLCTDCEPTSPQCVVARTLQKQIYMAIATKIALQMNCKGSELNRVDIPLKLYIV 543  
 QY 544 GIDCYHMTAGRRSIAAGFVASINEGTMFSCIFODRGGLVDGLKVCLOAALRAMNSC 603  
 DB 631 GIDCYHMTAGRRSIAAGFVASINEGTMFSCIFODRGGLVDGLKVCLOAALRAMNSC 690  
 QY 604 NEWPSRIIVYRDSVGGGLKTLVNYEVPOLDLCKSIGRGYNPLTVIVKRVNTRFF 663  
 DB 691 NEWPSRIIVYRDSVGGGLKTLVNYEVPOLDLCKSIGRGYNPLTVIVKRVNTRFF 750  
 QY 664 AOSGGRLONPDPGVIDVEYTRPEWYDFETVSOAVRSGSVSPTHYNYIDNSGLKPDHIQ 723  
 DB 751 AOSGGRLONPDPGVIDVEYTRPEWYDFETVSOAVRSGSVSPTHYNYIDNSGLKPDHIQ 810  
 QY 724 RLTYKLCIHYNNPVGIVRVPAPCOYAHKLAFLVQGSIHREPNLSLSNRLYYL 775  
 DB 811 RLTYKLCIHYNNPVGIVRVPAPCOYAHKLAFLVQGSIHREPNLSLSNRLYYL 862

## RESULT 6

QSNAG6 PRELIMINARY; PRT; 829 AA.

AC 09NAG6  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein FLJ35814 (P1wi).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Oshima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,  
 RA Arima M., Masehino K., Yuuki H., Hara H., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hiro Y., Saio K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamori K., Sekine M., Kikuchi H., Kanda K., Magatsuma M.,  
 RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,  
 RA Sugano S., Nagahari K., Masuno Y., Nagai K., Isega T.,  
 RT "MDO human cDNA sequencing project."  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK093133; BAC04068.1; -  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; P1W.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; P1wi; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; P1WI; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 829 AA; 94816 MW; 3AF9FD1D0FA9F5A7 CRC64;

Wed May 12 09:50:22 2004

us-10-043-774b-2.rpt

Page 5

Query Match  
Best Local Similarity 94.7%; Score 3883; DB 4; Length 829;  
Matches 733; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 4 GNTNQNLDHVKESKTSGGIIIVRLSTNHFRLTSRPPOMALYOYHIDYPLMEARLSAL 63
DB 90 GNTNQNLDHVKESKTSGGIIIVRLSTNHFRLTSRPPOMALYOYHIDYPLMEARLSAL 149
QY 64 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTINELPPTSPTCL 123
DB 150 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTINELPPTSPTCL 209
QY 124 QFYNIIFRRLKIKMLOOIGRNYNPNPIDPSHRLVIMPGFTTSILOYENSIMLCTDV 183
DB 210 QFYNIIFRRLKIKMLOOIGRNYNPNPIDPSHRLVIMPGFTTSILOYENSIMLCTDV 269
QY 184 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVVLTKYNNKTYRVDIDMDONPK 243
DB 270 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVVLTKYNNKTYRVDIDMDONPK 329
QY 244 STEFKADGSEVSEFLYYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPAMLIPELCYL 303
DB 330 STEFKADGSEVSEFLYYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPAMLIPELCYL 389
QY 304 TGLTTRKANDFVNMKDLAVHTRLTPEQORFVGRILIDYTHKDNVQRELAPMGISFDSNL 363
DB 390 TGLTTRKANDFVNMKDLAVHTRLTPEQORFVGRILIDYTHKDNVQRELAPMGISFDSNL 449
QY 364 LSFSGRILOTEKIHOGKTFDYNPOFADMSKETRGAPLISVYKLDNMLLITRANYEAN 423
DB 450 LSFSGRILOTEKIHOGKTFDYNPOFADMSKETRGAPLISVYKLDNMLLITRANYEAN 509
QY 424 SLIQNLKPYTPAMQOMKRAIMIEVDRTAVRLVLOQKVTADTOIVVCLLSNRKDXD 483
DB 510 SLIQNLKPYTPAMQOMKRAIMIEVDRTAVRLVLOQKVTADTOIVVCLLSNRKDXD 569
QY 484 AIKKYLCTDCCPTSPQCVVARTLGKOOTVMAITKIALQNMCMGSELNRVDIPKLKVMIV 543
DB 570 AIKKYLCTDCCPTSPQCVVARTLGKOOTVMAITKIALQNMCMGSELNRVDIPKLKVMIV 629
QY 544 GIDCYHDMTAGRRSIAGFVASINSGMTRFSCRIFQDRGQELVDGLKVCLOALPAMNSC 603
DB 630 GIDCYHDMTAGRRSIAGFVASINSGMTRFSCRIFQDRGQELVDGLKVCLOALPAMNSC 689
QY 604 NEYMSRIIVRDGVGDQGLKTLVNYEVQFLDKCSIGRGYNPLTVIIVKKRVNTRFF 663
DB 690 NEYMSRIIVRDGVGDQGLKTLVNYEVQFLDKCSIGRGYNPLTVIIVKKRVNTRFF 749
QY 664 AOSGRLQNPPLPGVIVDEVTRPEMYDDEFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 723
DB 750 AOSGRLQNPPLPGVIVDEVTRPEMYDDEFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 809
QY 724 RLTYKLCHMYNNWP 737
DB 810 RLTYKLCHMYNNWP 823
```

RESULT 7  
Q8UYX0 PRELIMINARY; PRT; 858 AA.  
AC Q8UYX0; 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_Taxid=7955;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Weeratane S.D., Gong Z., Tan C.-H.;  
RT "Cloning and characterization of zebrafish homolog of piwi, essential  
for germ-line stem cell self-renewal."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF36369; AAL57170.1; -  
DR InterPro; IPR003100; PAZ;  
DR InterPro; IPR003165; PIWI.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; PIWI; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; PIWI; 1.  
SQ SEQUENCE 858 AA; 97451 MW; 6A12F2511465777 CRC64;

Query Match  
Best Local Similarity 69.0%; Score 2828.5; DB 13; Length 858;  
Matches 509; Conservative 121; Mismatches 141; Indels 1; Gaps 1;

```
QY 4 GNTNQNLDHVKESKTSGGIIIVRLSTNHFRLTSRPPOMALYOYHIDYPLMEARLSAL 63
DB 88 GNTNQNLDHVKESKTSGGIIIVRLSTNHFRLTSRPPOMALYOYHIDYPLMEARLSAL 147
QY 64 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTINELPPTSPTCL 123
DB 148 LFQHEDLIGKCHAFDGTILFLPKRLQOKTEVEFSKTRNGEDVRIITLTINELPPTSPTCL 207
QY 124 QFYNIIFRRLKIKMLOOIGRNYNPNPIDPSHRLVIMPGFTTSILOYENSIMLCTDV 183
DB 208 QFYNIIFRRLKIKMLOOIGRNYNPNPIDPSHRLVIMPGFTTSILOYENSIMLCTDV 267
QY 184 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVVLTKYNNKTYRVDIDMDONPK 243
DB 268 SHKVLRSFVLDPMFNPHQTEHKEFOBOVSKEILGLVVLTKYNNKTYRVDIDMDONPK 327
QY 244 STEFKADGSEVSEFLYYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPAMLIPELCYL 303
DB 328 STEFKADGSEVSEFLYYRKQYNOEITDLKQPVLSQPKRRRGPGGTLPGPAMLIPELCYL 386
QY 328 NTFKRGD-TEISFKNYFKSQYGLDIDGNQVLLSHYKLGSPGRPPPGPAMLIPELCYL 366
DB 387 TGLTTRKANDFVNMKDLAVHTRLTPEQORFVGRILIDYTHKDNVQRELAPMGISFDSNL 446
QY 304 TGLTTRKANDFVNMKDLAVHTRLTPEQORFVGRILIDYTHKDNVQRELAPMGISFDSNL 363
DB 387 TGLTTRKANDFVNMKDLAVHTRLTPEQORFVGRILIDYTHKDNVQRELAPMGISFDSNL 446
QY 364 LSFSGRILOTEKIHOGKTFDYNPOFADMSKETRGAPLISVYKLDNMLLITRANYEAN 423
DB 447 LSFSGRILOTEKIHOGKTFDYNPOFADMSKETRGAPLISVYKLDNMLLITRANYEAN 506
QY 424 SLIQNLKPYTPAMQOMKRAIMIEVDRTAVRLVLOQKVTADTOIVVCLLSNRKDXD 483
DB 507 SLIQNLKPYTPAMQOMKRAIMIEVDRTAVRLVLOQKVTADTOIVVCLLSNRKDXD 566
QY 484 AIKKYLCTDCCPTSPQCVVARTLGKOOTVMAITKIALQNMCMGSELNRVDIPKLKVMIV 543
DB 567 AIKKYLCTDCCPTSPQCVVARTLGKOOTVMAITKIALQNMCMGSELNRVDIPKLKVMIV 626
QY 544 GIDCYHDMTAGRRSIAGFVASINSGMTRFSCRIFQDRGQELVDGLKVCLOALPAMNSC 603
DB 627 GIDCYHDMTAGRRSIAGFVASINSGMTRFSCRIFQDRGQELVDGLKVCLOALPAMNSC 686
QY 604 NEYMSRIIVRDGVGDQGLKTLVNYEVQFLDKCSIGRGYNPLTVIIVKKRVNTRFF 663
DB 687 NEYMSRIIVRDGVGDQGLKTLVNYEVQFLDKCSIGRGYNPLTVIIVKKRVNTRFF 746
QY 664 AOSGRLQNPPLPGVIVDEVTRPEMYDDEFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 723
DB 747 AOSGRLQNPPLPGVIVDEVTRPEMYDDEFIVSQAVRSQVSPTHYVNIYDNGSLKPDHIQ 806
QY 724 RLTYKLCHMYNNWP 737
DB 807 RLTYKLCHMYNNWP 823
```

RESULT 8  
Q72323 PRELIMINARY; PRT; 882 AA.



AC Q7223; 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE P1WIL3 protein.  
 GN P1WIL3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBT\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22788786; PubMed=12906857;  
 RA Sasaki T., Shiohama A., Minoshima S., Shimizu N.,  
 RT "Identification of eight members of the Argonaute family in the human  
 genome.";  
 RT Genomics 82:323-330(2003).  
 DR EMBL, AB079366; BAC01343.1; -  
 SQ SEQUENCE 882 AA; 101121 MW; BD65CA74C3D36DB8 CRC64;

Query Match 52.3%; Score 2144; DB 4; Length 882;  
 Best Local Similarity 52.0%; Pred. No. 2.6e-162;  
 Matches 407; Conservative 139; Mismatches 222; Indels 14; Gaps 5;

QY 5 VNTKRONLDHVKESTKSSGIIIVRLSTNFRILTSRPOALYOHYHIDNPLMEARRLSALL 64  
 DB 104 VNTKRONLDHVKESTKSSGIIIVRLSTNFRILTSRPOALYOHYHIDNPLMEARRLSALL 163  
 QY 65 FCHEDLIGKCAFDGTLIFLPRLOQKTEVFSKTRNEDVRIITLNNELPPTSPCLQ 124  
 DB 164 DQHRKPKERRHIFDONSLLSRPLKERVLELSTTKDNVYKIVLSESKELTPTSPCLR 223  
 QY 125 FYNIFRRLKIMNLOIGRNYNPNNDIDIPSH--RLVTPGTTSLIOYENSIMCTD 182  
 DB 224 YNIFRRLKIMNLOIGRNYNPNNDIDIPSH--RLVTPGTTSLIOYENSIMCTD 283  
 QY 183 VSHKVLRSSEYLDLDFENFYHOTEEHKFOEVSKEILGLVLTAKNNKYRVDIDMDQNP 242  
 DB 284 VSHKVLRSSEYLDLDFENFYHOTEEHKFOEVSKEILGLVLTAKNNKYRVDIDMDQNP 343  
 QY 243 KSTFKKADGSEVSLEYRYKQYNGETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPELCY 302  
 DB 344 ETEFKSDGSKITVYIDYRQOQKEIVYKQPLVYSCGRWKGLGTGOREBILIPDICH 403  
 QY 303 LTGLTDXKRNPNWVKDLAVHTRLPBQOREVGLIDYHKNNVORELSDWGLSPDSN 362  
 DB 404 MTGLTDXKRNPNWVKDLAVHTRLPBQOREVGLIDYHKNNVORELSDWGLSPDSN 463  
 QY 463 LLSFSGRILOTEKIHOGKTFDYNFOFADWSKETGAPLISVKPLDNLIIYTRNRYEAA 422  
 DB 522 FLTSVGRVILKXNITVQGRRWYKNSQ--GDWSREIREFLNAMPILSLIYSSRSREA 522  
 QY 423 NSLIONLREKVPAMQOKRKAIMEVDRTETAYLRVLOQKTYADTQI-----VYCL 473  
 DB 522 NSLIONLREKVPAMQOKRKAIMEVDRTETAYLRVLOQKTYADTQI-----VYCL 582  
 QY 582 MSLKHLQSVTAEMGIVTKPAEMIEVDSDANSYIDTLTKYRPTLQKMSCLLYFKYICI 582  
 DB 582 MSLKHLQSVTAEMGIVTKPAEMIEVDSDANSYIDTLTKYRPTLQKMSCLLYFKYICI 641  
 QY 474 LLSNRKDKYAIKXVLTCTDCTPSQCVARTLQKQYMAIATKIALONNCKMGEELMRV 533  
 DB 641 LLSNRKDKYAIKXVLTCTDCTPSQCVARTLQKQYMAIATKIALONNCKMGEELMRV 641  
 QY 534 DIPKLAVTGVIGDCHDMTAGRRSIAGYASINSGMTWFSRCIFQDNGOGLINDGLKYL 593  
 DB 642 DIPKLAVTGVIGDCHDMTAGRRSIAGYASINSGMTWFSRCIFQDNGOGLINDGLKYL 701  
 QY 642 ETDVORTMTVEGIDCHDMTAGRRSIAGYASINSGMTWFSRCIFQDNGOGLINDGLKYL 701  
 DB 642 ETDVORTMTVEGIDCHDMTAGRRSIAGYASINSGMTWFSRCIFQDNGOGLINDGLKYL 760  
 QY 594 QALDAMNSCNEVPSRIIVYRDGVGQGLKTLVYEVPOFADCKLSIGRGVNPFLTVIV 653  
 DB 760 QALDAMNSCNEVPSRIIVYRDGVGQGLKTLVYEVPOFADCKLSIGRGVNPFLTVIV 760  
 QY 702 KALDAMNSCNEVPSRIIVYRDGVGQGLKTLVYEVPOFADCKLSIGRGVNPFLTVIV 760  
 DB 702 KALDAMNSCNEVPSRIIVYRDGVGQGLKTLVYEVPOFADCKLSIGRGVNPFLTVIV 820  
 QY 654 VKKRVNTRFPAQSGRLQNLPLFGYIVDYEVTREPMYDFFIYSGVARSQSVSPTRHNYVD 713  
 DB 820 VKKRVNTRFPAQSGRLQNLPLFGYIVDYEVTREPMYDFFIYSGVARSQSVSPTRHNYVD 820  
 QY 761 VKKRVNTRFPAQSGRLQNLPLFGYIVDYEVTREPMYDFFIYSGVARSQSVSPTRHNYVD 820  
 DB 820 VKKRVNTRFPAQSGRLQNLPLFGYIVDYEVTREPMYDFFIYSGVARSQSVSPTRHNYVD 820  
 QY 714 NSGLKPDHQLRLTYLCHLYYVWPGVIRVAPCAQYAKLAFVWGQISREPRILSNRLY 773  
 DB 773 NSGLKPDHQLRLTYLCHLYYVWPGVIRVAPCAQYAKLAFVWGQISREPRILSNRLY 773

DB 821 TIGLSPDVTQRLTYLCHLYYVWPGVIRVAPCAQYAKLAFVWGQISREPRILSNRLY 880  
 QY 774 YL 775  
 DB 881 YL 882

RESULT 9  
 Q8NEH2 PRELIMINARY; PRI; 852 AA.  
 AC Q8NEH2; 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Similar to p1w1 like homolog 1 (Drosophila).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBT\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL, BC031060; AAH31060.1; -  
 DR InterPro: IPR003100; PAZ.  
 DR InterPro: IPR003165; P1W1.  
 DR Pfam: PF02170; PAZ; 1.  
 DR Pfam: PF02170; P1W1; 1.  
 DR PROSITE: PS00821; PAZ; 1.  
 DR PROSITE: PS00822; P1W1; 1.  
 SQ SEQUENCE 852 AA; 96588 MW; 37769BE078B96D13 CRC64;

Query Match 51.3%; Score 2103.5; DB 4; Length 852;  
 Best Local Similarity 51.2%; Pred. No. 4.2e-159;  
 Matches 398; Conservative 146; Mismatches 225; Indels 9; Gaps 4;

QY 1 MIFGVNTRONLDHVKESTKSSGIIIVRLSTNFRILTSRPOALYOHYHIDNPLMEARRLS 60  
 DB 81 MIFGVNTRONLDHVKESTKSSGIIIVRLSTNFRILTSRPOALYOHYHIDNPLMEARRLS 140  
 QY 61 SALLFOHEDLIGKCAFDGTLIFLPRLOQKTEVFSKTRNEDVRIITLNNELPPTSP 120  
 DB 141 SALLFOHEDLIGKCAFDGTLIFLPRLOQKTEVFSKTRNEDVRIITLNNELPPTSP 200  
 QY 121 TCIQFNIIIFRRLKIMNLOIGRNYNPNNDIDIPSHRLVMPGTTSLIOYENSIMCTD 180  
 DB 201 TCIQFNIIIFRRLKIMNLOIGRNYNPNNDIDIPSHRLVMPGTTSLIOYENSIMCTD 260  
 QY 201 VCIQFNIIIFRRLKIMNLOIGRNYNPNNDIDIPSHRLVMPGTTSLIOYENSIMCTD 260  
 DB 261 VCIQFNIIIFRRLKIMNLOIGRNYNPNNDIDIPSHRLVMPGTTSLIOYENSIMCTD 320  
 QY 181 TDVSHKVLRSSEYLDLDFENFYHOTEEHKFOEVSKEILGLVLTAKNNKYRVDIDMDQ 240  
 DB 320 TDVSHKVLRSSEYLDLDFENFYHOTEEHKFOEVSKEILGLVLTAKNNKYRVDIDMDQ 320  
 QY 241 NPSTFKKADGSEVSLEYRYKQYNGETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPEL 300  
 DB 321 NPSTFKKADGSEVSLEYRYKQYNGETIDLPKOPVLSQPKRRRPGGTLPGPAMLIPEL 379  
 QY 301 KYTHITQKBDGHEITVVDYKQYDITVSDLNQPMVLSLKKKRNDSNAG--LAHLIPBL 379  
 DB 380 KYTHITQKBDGHEITVVDYKQYDITVSDLNQPMVLSLKKKRNDSNAG--LAHLIPBL 439  
 QY 380 CFLTGLTDQATSDFOKAKVAIAKTRLSFGRCQRLRLVDNIOQNTNARFELTGWLFHG 439  
 DB 439 CFLTGLTDQATSDFOKAKVAIAKTRLSFGRCQRLRLVDNIOQNTNARFELTGWLFHG 494  
 QY 361 SNLISFSGRILOTEKIHOGKTFDYNFOFADWSKETGAPLISVKPLDNLIIYTRNRYEAA 417  
 DB 494 SNLISFSGRILOTEKIHOGKTFDYNFOFADWSKETGAPLISVKPLDNLIIYTRNRYEAA 494  
 QY 440 SQ--ISLGRVPEKILMQ---DHTCQVSAABSDKDRITCKIILAQNSLNTLILCSBR 494  
 DB 494 SQ--ISLGRVPEKILMQ---DHTCQVSAABSDKDRITCKIILAQNSLNTLILCSBR 554  
 QY 418 NYEANSLLIONLREKVPAMQOKRKAIMEVDRTETAYLRVLOQKTYADTQIYVCLSSN 477  
 DB 554 NYEANSLLIONLREKVPAMQOKRKAIMEVDRTETAYLRVLOQKTYADTQIYVCLSSN 554  
 QY 478 RKQKDAIKKYLCTDCTPSQCVARTLQKQYMAIATKIALONNCKMGEELMRV 537  
 DB 554 RKQKDAIKKYLCTDCTPSQCVARTLQKQYMAIATKIALONNCKMGEELMRV 537  
 QY 555 QKTYDYSIKKYLSSDCDVPBQCVLARTLNKQGMKMSIATKIALQKTKGGEELMAVEIRP 614  
 DB 614 QKTYDYSIKKYLSSDCDVPBQCVLARTLNKQGMKMSIATKIALQKTKGGEELMAVEIRP 614

QY 538 KLWMTGIDCYHDMTAGRSIAGFVASINEGNTWFSRCIFQDRQGLVDGLKYCLOAL 597  
 DB 615 KSLMVGIDVCKDALSCKVMVGVASVNPRIITWFSRCILQRTWTDVADCLKFMGAL 674  
 QY 598 RAMNSCNEYMPRIIYRDGVGDQGLKTLVNEVQFIDCLKSIGRGVNPETIYVKKR 657  
 DB 675 NKMYKNHDLPRILIIYRAGVGDQGLKTLIEVEVQFLSSVASSNTSRSLVIVRKK 734  
 QY 658 VNTREFAOSGGRLONPFGTIVDEVTREPWTDFEIVSQAVRSGVSPTHYNYTNSGL 717  
 DB 735 CMPEFTEEMKRTVQNPPIGLTVDSBATRNEVDFLLISQVACRGTVSPTYNYVTDNGL 794  
 QY 718 KPDHQRLLTYLCHLYYNNPGVIRVAPCOYAHKLAFLVGSISIREPNTLSNRLYYL 775  
 DB 795 KPDHQRLLTYLCHLYYNNPGVIRVAPCOYAHKLAFLVGSISIREPNTLSNRLYYL 852

## RESULT 10

Q8N9V8 PRELIMINARY; PRT; 852 AA.

AC 08N9V8; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein FLJ36156 (p1w1).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Testis;  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Hoshida T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Magatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;  
 RT "NEO human cDNA sequencing project."  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK093475; BAC04179.1; -  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; P1w1.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; P1w1; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; P1w1; 1.  
 DR KW Hypothetical protein.  
 SQ SEQUENCE 852 AA; 96560 MW; 75C6FEBAE70701B CRC64;

Query Match 51.2%; Score 2101.5; DB 4; Length 852;  
 Best Local Similarity 51.0%; Pred. No. 61e-159;  
 Matches 397; Conservative 148; Mismatches 224; Indels 9; Gaps 4;

QY 1 MIFGVNTRONLDHVESKTSKSSGIIIVRLSTNHRFLTSRPOALYOHIDYNPMEARRLR 60  
 DB 81 MDLSICTREKLAHVNRCKAGSSGIPVKLVTLNFDLPQDQOLYOHVTVYFDLASRRLR 140  
 QY 61 SALLFQHEDLIGKCAFPDGTILFLPKLQCKYTEVFSKTRNGEDVRLITLTNLEPPTSP 120  
 DB 141 IALLVSHSELSENKARAFDGAIFLFSQKLEKTELESSEFQRETTIKMTTLKRELPSSSP 200  
 QY 121 TCLQFYNIIFRLRLKIMNLOQIGRNYNPNPDIIPSHRLVIMPGETTSLQYENSIMTC 180  
 DB 201 VCIOGFNIIFRLKILKSMYQIGRNFYNPSPMEIPQKLSLWPGFAISVGFPERKILFS 260  
 QY 161 TDVSHKVLRSFTVDFMNFYHQTBEHFKQEQVSKELIGLVLTNKANKTKRVVDIDWQ 240  
 DB 261 ADVSKVLRNFTVLEFMTALCQRTGLSCFTQCEKQILGLIVLRNNRYSIDIDIMSV 320

QY 241 NPKSTFKADGSEVSLEYRYKQYNOETDLPKQVILVSPKRRRGCGTLPEPAMILPEL 300  
 DB 321 KPTHFOKDGDTETIYVDYKQYDITVSDINQPMILVSLKKRRNNSDAQ-LAHILPEL 379  
 QY 301 CYLTGLTDKRNDFNVWKLAVHTRILTPROGOREVRLIDYIHKNDNVORELMDGLSPD 360  
 DB 380 CFLTGLTDQATSPQMLKAVAEKTRLSPSGRQRLARLVNDIQRNNAFLEETWGLHFG 439  
 QY 361 SNLSPSGRILOTEKIHGGKTFDYNPQ--PADMSKETRGAPLISVKELDWLLIYTRR 417  
 DB 440 SQ-LSLTGRIVPSEKILMO---DHICQPVASADWSKDRCTCKILNAQSLMTWLILCSOR 494  
 QY 418 NYEANSILQNFKYTPAMGMOKRAKIMTEVDRTAVALRYVQOKTADTOIVVCLLSN 477  
 DB 495 TEYVASEFLNCLFRVAGSMGFNDYPKIKVGENPAAFYRAIQYVDPVQVLMCLPSN 554  
 QY 478 RKDKYDAIKKYLCTDPTSPQCVARTLGGKQOTVMAIATKIALQNNCKMGGELMRVDIP 537  
 DB 555 OKTYYSIKKYLSSDQVPSPQCVARTLGGKQOTVMAIATKIALQNNCKMGGELMRVDIP 614  
 QY 538 KLWMTGIDCYHDMTAGRSIAGFVASINEGNTWFSRCIFQDRQGLVDGLKYCLOAL 597  
 DB 615 KSLMVGIDVCKDALSCKVMVGVASVNPRIITWFSRCILQRTWTDVADCLKFMGAL 674  
 QY 598 RAMNSCNEYMPRIIYRDGVGDQGLKTLVNEVQFIDCLKSIGRGVNPETIYVKKR 657  
 DB 675 NKMYKNHDLPRILIIYRAGVGDQGLKTLIEVEVQFLSSVASSNTSRSLVIVRKK 734  
 QY 658 VNTREFAOSGGRLONPFGTIVDEVTREPWTDFEIVSQAVRSGVSPTHYNYTNSGL 717  
 DB 735 CMPEFTEEMKRTVQNPPIGLTVDSBATRNEVDFLLISQVACRGTVSPTYNYVTDNGL 794  
 QY 718 KPDHQRLLTYLCHLYYNNPGVIRVAPCOYAHKLAFLVGSISIREPNTLSNRLYYL 775  
 DB 795 KPDHQRLLTYLCHLYYNNPGVIRVAPCOYAHKLAFLVGSISIREPNTLSNRLYYL 852

## RESULT 11

Q72324 PRELIMINARY; PRT; 852 AA.

AC Q72324; 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE P1W1L1 protein.  
 GN P1W1L1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22788786; PubMed=12906857;  
 RA Sasaki T., Shiohama S., Shimizu N.;  
 RT "Identification of eight members of the Argonaute family in the human  
 genome."  
 RL Genomes 82:323-330 (2003).  
 DR EMBL; AB079366; BAC81341.1; -  
 SQ SEQUENCE 852 AA; 96558 MW; 3777A264A2114D56 CRC64;

Query Match 51.2%; Score 2098.5; DB 4; Length 852;  
 Best Local Similarity 51.0%; Pred. No. 1.1e-158;  
 Matches 397; Conservative 146; Mismatches 226; Indels 9; Gaps 4;

QY 1 MIFGVNTRONLDHVESKTSKSSGIIIVRLSTNHRFLTSRPOALYOHIDYNPMEARRLR 60  
 DB 81 MDLSICTREKLAHVNRCKAGSSGIPVKLVTLNFDLPQDQOLYOHVTVYFDLASRRLR 140  
 QY 61 SALLFQHEDLIGKCAFPDGTILFLPKLQCKYTEVFSKTRNGEDVRLITLTNLEPPTSP 120  
 DB 141 IALLVSHSELSENKARAFDGAIFLFSQKLEKTELESSEFQRETTIKMTTLKRELPSSSP 200  
 QY 121 TCLQFYNIIFRLRLKIMNLOQIGRNYNPNPDIIPSHRLVIMPGETTSLQYENSIMTC 180

[illegible]

	Best Local Similarity 50.5%; Pred. No. 1,46-151; Matches 386; Conservative 133; Mismatches 226; Indels 9; Gaps 6
QY	14 VRESKTSGGSGITVRLSTHFRFLTSRPOKALYOHVIDNPMLMEARRLSALLFOHEDLIGK 73
Db	14 VRESKTSGGSGITVRLSTHFRFLTSRPOKALYOHVIDNPMLMEARRLSALLFOHEDLIGK 73
QY	98 VYQALAGDK--IALIANGFKLTKRPMQOLQYVDEPEPLINRAFPALKGSHALLGK 154
Db	98 VYQALAGDK--IALIANGFKLTKRPMQOLQYVDEPEPLINRAFPALKGSHALLGK 154
QY	74 CHAFDGLTILFLPKRLQCKTEVFSKTRNGEDVRLTITLFTNELPSTSPCQFYIIFRRL 133
Db	74 CHAFDGLTILFLPKRLQCKTEVFSKTRNGEDVRLTITLFTNELPSTSPCQFYIIFRRL 133
QY	134 LKIMNIQOIGRNYNPNDPIDPSHRLVYAFGFTSLILOENSLMLCTDVSRYKLBSFTY 193
Db	134 LKIMNIQOIGRNYNPNDPIDPSHRLVYAFGFTSLILOENSLMLCTDVSRYKLBSFTY 193
QY	215 LKIMINNEQGRNYVPTPAIDIKOHGQLOMPGFSTILQYEYDMLSDISHKYLRLQTV 274
Db	215 LKIMINNEQGRNYVPTPAIDIKOHGQLOMPGFSTILQYEYDMLSDISHKYLRLQTV 274
QY	194 LDPMEFYHQTEHEHFOEVSKEILGLVLLFRKNNKYRVDDIDMDQNPSTFEKADGSE 253
Db	194 LDPMEFYHQTEHEHFOEVSKEILGLVLLFRKNNKYRVDDIDMDQNPSTFEKADGSE 253
QY	275 WEYMNDLPKKA-RGPFKEITKMLGCVLTKNNKYRVDDIDFDTPADTFETRSG-P 332
Db	275 WEYMNDLPKKA-RGPFKEITKMLGCVLTKNNKYRVDDIDFDTPADTFETRSG-P 332
QY	254 VSFELEYRKQNOEITDLOKQVLYSQPKRRRPGGTLPGPAMLPELCYLTGLTDKRRND 313
Db	254 VSFELEYRKQNOEITDLOKQVLYSQPKRRRPGGTLPGPAMLPELCYLTGLTDKRRND 313
QY	333 VSVVDYFKKSYRVLHDVNOQPMVSRPFKREKKGV--GPAYVLPPELCFTGLSDDDRAD 350
Db	333 VSVVDYFKKSYRVLHDVNOQPMVSRPFKREKKGV--GPAYVLPPELCFTGLSDDDRAD 350
QY	314 FNNMVDLANHTELTPEQGRQREGRRLIDVYHNKDVQSELRDMGSPDSNLSFSGRLQT 373
Db	314 FNNMVDLANHTELTPEQGRQREGRRLIDVYHNKDVQSELRDMGSPDSNLSFSGRLQT 373
QY	391 FNNMCKMGHTVGVQDRCRTLSGFTKKLSNEEKYTLDSMGWEPFKEQVLTGRVLP? 450
Db	391 FNNMCKMGHTVGVQDRCRTLSGFTKKLSNEEKYTLDSMGWEPFKEQVLTGRVLP? 450
QY	374 EKHQGGKTFDYNPOFAPWSKETRGAPLISYKPLDNMLLYTRRNVAANSLLQNLKVT 433
Db	374 EKHQGGKTFDYNPOFAPWSKETRGAPLISYKPLDNMLLYTRRNVAANSLLQNLKVT 433
QY	451 EKLFGQGRQFSYNPENADWSRDRGNALTDKILNNMKIFYTRDANRGODFIKSLVRA 510
Db	451 EKLFGQGRQFSYNPENADWSRDRGNALTDKILNNMKIFYTRDANRGODFIKSLVRA 510
QY	434 PAKGQMRKALITEV-DRTEAYLYAVLQCKYTADTQIVYCLSSNRDKDAIKKYICTD 497
Db	434 PAKGQMRKALITEV-DRTEAYLYAVLQCKYTADTQIVYCLSSNRDKDAIKKYICTD 497
QY	511 NPGNNVNRBEIVELPDRTEETRYRSLOAQIAQOTQIVAVLLPNNRDRDPAIKKTCVTT 570
Db	511 NPGNNVNRBEIVELPDRTEETRYRSLOAQIAQOTQIVAVLLPNNRDRDPAIKKTCVTT 570
QY	493 CPTSPSCVARTILGKQQTVMALATKIALQNNCKKMGSELMRVDIPLKVMTVGIDCYHMT 552
Db	493 CPTSPSCVARTILGKQQTVMALATKIALQNNCKKMGSELMRVDIPLKVMTVGIDCYHMT 552
QY	571 HPPSPQVLYSRILSKQQLMMSVATKIAMQNNCKKMGSGDLMRVEIPLSNMLIIGDSDYHSL 630
Db	571 HPPSPQVLYSRILSKQQLMMSVATKIAMQNNCKKMGSGDLMRVEIPLSNMLIIGDSDYHSL 630
QY	553 AGRSIAGSVASINEGMRWPSRCIFPDQGGELVDGLKVCQALAPRANSCNEMPSRIT 612
Db	553 AGRSIAGSVASINEGMRWPSRCIFPDQGGELVDGLKVCQALAPRANSCNEMPSRIT 612
QY	631 TKGRSVGLGVASAMNKSQTSFSSCAFOHAGEFGANLSTLMNMLKRYQINEXFPERII 690
Db	631 TKGRSVGLGVASAMNKSQTSFSSCAFOHAGEFGANLSTLMNMLKRYQINEXFPERII 690
QY	613 VYADVGDDQQLTLVNVYEPQGLDCT-SSIRGYNPRLTVYVKKRVNTRFFAQSGLRQ 671
Db	613 VYADVGDDQQLTLVNVYEPQGLDCT-SSIRGYNPRLTVYVKKRVNTRFFAQSGLRQ 671
QY	691 IFPDGSDGQVLLVVDYELKQIKDILDKYQGYGHKLAIVVYKGRINNRFPALRGGLS 750
Db	691 IFPDGSDGQVLLVVDYELKQIKDILDKYQGYGHKLAIVVYKGRINNRFPALRGGLS 750
QY	672 NPLPGTVIDVENTRPEWYDFPIVSGAVSAGSVSPTHNVYIDNSGLKPDHIGRLTYLCH 731
Db	672 NPLPGTVIDVENTRPEWYDFPIVSGAVSAGSVSPTHNVYIDNSGLKPDHIGRLTYLCH 731
QY	732 IYNNPGVIRVAPQCYAHKALFVLSGSIHREPILSNRLXYL 775
Db	732 IYNNPGVIRVAPQCYAHKALFVLSGSIHREPILSNRLXYL 775
QY	811 LYNNMGTVRVAPCMYAHKALFVLSGSIHREPILSNRLXYL 854
Db	811 LYNNMGTVRVAPCMYAHKALFVLSGSIHREPILSNRLXYL 854
RESULT 13	
08CGT6 PRELIMINARY: PRT: 878 AA.	
AC 08CGT6; 01-MAR-2003 (Tremblrel. 23, Created)	
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)	
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)	
DE MW1.2 protein.	
GN MW12.	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=22301763; PubMed=12414724;	
ZA Carmell M.A., Xuan Z., Zhang M.Q., Hannon G.J.;	

"The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis."

RT Genes Dev. 16:2733-2742 (2002).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Adams M., Mural R.,  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AY155692; AAN75583.1; --  
 DR InterPro; IPR003100; PAZ.  
 DR InterPro; IPR003165; P1w1.  
 DR Pfam; PF02170; PAZ; 1.  
 DR Pfam; PF02171; P1w1; 1.  
 DR PROSITE; PS50821; PAZ; 1.  
 DR PROSITE; PS50822; P1w1; 1.  
 SQ SEQUENCE 878 AA; 96617 MW; 3C6265E0E4ED7C3F CRC64;

Query Match 45.3%; Score 1859.5; DB 11; Length 878;  
 Best Local Similarity 46.5%; Pred. No. 1.5e-139;  
 Matches 370; Conservative 134; Mismatches 209; Indels 83; Gaps 10;

4 GVNTRNDHYKESKTSGGIIVRLSTNHPRLSRPQWALYQYHIDYNPIMLEARRLSAL 63  
 DB 142 GVTREKLTTHVXDCTGSSGIPVRLVTNLEFMDLPQDMQLYQHVTYSPDLASRRRLAL 201  
 64 LFEHEDLIGKCHAPDGTLLPLPKRLQOKVTEVPSKTRNGEDVRIITITLNEPPTPTCL 123  
 DB 202 LYNHSLISDKAKAFDGLFLSEKLDQKVELTSETQGETGIKITLTLLTSKLFENSPVCI 261  
 124 QFNVTIRRLKIKMLQOIGRNYNPNPDIDIPSHLVMEGFTSLTLOENSMICTDV 183  
 DB 262 QFNVTIRRLKIKMLQOIGRNYNPNPDIDIPSHLVMEGFTSLTLOENSMICTDV 183  
 184 SHKVLRESEVLDVEMFNHYQTEHKFOVSKELIGLVLLTKNNKTYRVDIDMDQNF 243  
 DB 306 NYKTLRETVLDFMTDLCTRGMSCFEMCHKQVLGLVLLTRNNKTYRIDIDMSVKFT 365  
 244 STPKADGSEVSELYRKOYNOETIDKQPVLVSQPRRRGPGGTLPPAMLIPELCYL 303  
 DB 366 QAFQKRGSEVTVYDYVYQYDITLSDLNQPVLSLKRKNDNSE-PQWVHLMPELCFL 424  
 304 TGLDKNRNDPNWMDLAVHTRLTPEOROREVGLIDYIHNDVORELND-----W-- 355  
 DB 425 TGLSSQATSDPRMLKAVAEETRLSPVGRQQLAVLDIGRTLPSSGEVLSHSLPIMAP 484  
 356 ---GLSPD---SNLSFSGRILOTEKIHOG-----GKTFDYNPQF-ADMSKETEGA 399  
 DB 485 EPGGLSSAIPSTVLPFAQQLTLTSLSPGIPLPLKPSFLFCQPAFAADMSKDWSC 544  
 400 PLISVPELDNMLLYTRNRYEANSLSLONLFKTPAMQMOKRAIMIEVDRTAAYRLV 459  
 DB 545 KVLSSQPLNRLIYVCNNAEHLIEAFSLCRRVSGSGENGVKXIKIKVDTPAFAKAI 604  
 460 QOKTADTQIVVCLLSNRKDKYDAIKKYLCTDCTPSPQCVAVATLKKOQTVAIAIKIA 519  
 DB 605 QVHGDPRQVLMVCLIPSNQKYVDSIKKYLSSDCFPVSOQVLTRTLKNGQMLSVATKIA 664  
 520 LQNMCKKGGELMRVDIPLKLVMIYIGIDCYHMTAGRSIAGFVASINEGTRMFSRCIFQ 579  
 DB 665 MQMCKKGGELMRVDIPLKLVMIYIGIDCYHMTAGRSIAGFVASINEGTRMFSRCIFQ 579  
 580 DRGELVGLKVCQALRAWNSCNEVPSRIIYRDGVGQGLKTVNVEPFLDCLX 639  
 DB 725 RTADADLCKVCMTGALNRYRHHNDIPARIIVYRGVNGQKAVLEAVEVPLLSVY 784  
 640 SIGRGINRLTVIYVKKRVNTRFPAQSGRLQNPFGVIDVETVREBMDFFVSGAVR 699  
 DB 785 ECG-----SDAR-----DFFYLIISOTAN 802  
 700 SSGVSPTHYNYIYDNGSLKPDHIOQLTYKCHIIYNNPFGVIRVAPQYAKLAFVLGQS 759  
 DB 803 RGTYSPTHYNYIYDNGSLKPDHIOQLTYKCHIIYNNPFGVIRVAPQYAKLAFVLGQS 759  
 760 IHREPNLSLNRLLYYL 775

DB 863 VAKPSLELANLRYL 878

RESULT 14

ID 08N69

PRELIMINARY;

PRT; 666 AA.

AC 08N69

01-OCT-2002 (TRENBLREL. 22, Created)

DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)

01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE Hypothetical protein Flj39518 (p1w1).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_Taxid=9606;

RA SEQUENCE FROM N.A.

Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,

RA Tanigami A., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,

Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,

RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,

Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Watanabe M., Murakawa K., Kaneshiro K., Takahashi-Fuji A., Oshima A.,

Masuhara A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,

RA Masuhara A., Nagai K., Isogai T.,

"NEO human cDNA sequencing project."

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

EMBL; AK096837; BAC04873.1; --

DR InterPro; IPR003100; PAZ.

InterPro; IPR003165; P1w1.

DR Pfam; PF02170; PAZ; 1.

Pfam; PF02171; P1w1; 1.

DR PROSITE; PS50821; PAZ; 1.

PROSITE; PS50822; P1w1; 1.

KW Hypothetical protein.

SEQUENCE 666 AA; 76039 MW; 1082D86916390FA8 CRC64;

Query Match 44.8%; Score 1838.5; DB 4; Length 666;  
 Best Local Similarity 51.5%; Pred. No. 4.5e-138;  
 Matches 346; Conservative 127; Mismatches 190; Indels 9; Gaps 4;

107 ITTTLNLPPTSPCTQPFYNIIPRLKIKMLQOIGRNYNPNPDIDIPSHLVMEGFT 166  
 DB 1 MTTLKRELPSSSPCIQVFNIIIRKILKLSMYQIGRNFYPSPEMELPQKLSIMPEF 60  
 167 TTSILOENSMICTDVSHKVLRESEVLDVEMFNHYQTEHKFOVSKELIGLVLLTKY 226  
 DB 61 AIVSYFEKKLFSDADSVYKVLRNETHLEFMTALCQRTGLSCFTQCEQQLGLVLTLY 120  
 227 NNTYRVDIDMDQNPXSTFKKADGSEVSELYRKOYNOETIDKQPVLVSQPRRRGPG 286  
 DB 121 NNTYRVDIDMDQNPXSTFKKADGSEVSELYRKOYNOETIDKQPVLVSQPRRRGPG 286  
 287 GGLTPGAMLIPELCYLNLGIDKMRNDFWMDLAVHTRLTPEOROREVGLIDYIHND 346  
 DB 181 NSE-PQLAHILPELCYLTGLTQATSDFQMKAVAEKTRLSGQQRILARLVNDIQNT 239  
 347 NQRELRDGLSPDSNLSFSGRILOTEKIHOGKTFDYNPQ---FADMSKETEGAFLIS 403  
 DB 240 NARFELTEGLFFGSG-ISTGRIVSEKILMQ---DHICQPVSAADMSKQIRTKIIN 294  
 404 VPELDNMLLYTRNRYEANSLSLONLFKTPAMQMOKRAIMIEVDRTAAYRLVLOQY 463  
 DB 295 AQSINWMLTLCGRTEYVAESFLNCRARYTGSMGVNDVPKLIKQENPAFAFPAQYV 354  
 464 TADTQIVVCLLSNRKDKYDAIKKYLCTDCTPSPQCVAVATLKKOQTVAIAIKIA 523  
 DB 355 DDPVQVLMVCLIPSNQKYVDSIKKYLSSDCFPVSOQVLTRTLKNGQMLSVATKIA 414  
 524 CKMGELMRVDIPLKLVMIYIGIDCYHMTAGRSIAGFVASINEGTRMFSRCIFQDRQ 583  
 DB 415 CULGSELMAVEIPLKLSIMVGTIDVCDALSKQVMVYGVASVNPRTTRFMSRCIIQRTW 474

Wed May 12 09:50:22 2004

us-10-043-774b-2.rspc

Page 10

QY 584 ELVDGKVCLOALBANS CNEYMPRIIVYRDVGDLKTLVNEVDFLDCLKSIGR 643  
DB 475 DVADCLKVFETGALNKYKNNHDLPAIIYRAGVGGDLKTLIEVEVQLSSVABSSS 534  
QY 644 GINDELTVIVKRVKNTREFFAOGGRLONPLGTVIDVETREPTVDFIVSQAVSSGV 703  
DB 535 NTSRSLSVIVKRVKNTREFFAOGGRLONPLGTVIDVETREPTVDFIVSQAVSSGV 594  
QY 704 SPTNYIVYNSGLKPHIOPLTYKLCIYNNMPGVIRVAPQCYAHKLAFLVQSIHRE 763  
DB 595 SPTNYIVYNSGLKPHIOPLTYKLCIYNNMPGVIRVAPQCYAHKLAFLVQSIHRE 654  
QY 764 PNLISLNRLLYL 775  
DB 655 PSLBLANHLFYL 666

RESULT 15  
QYGPAT7 PRELIMINARY; PRT; 808 AA.  
AC QYGPAT7;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Seawi (Fragment).  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;  
OC Strongylocentrotus.  
CX NCBI\_TaxID=7668;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rodriguez A.J., Bonder E.M.;  
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin  
RT Homolog of Pw1."  
RL Submitted (NCV-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY014900; AAG42534.1; -  
DR InterPro; IPR003100; PAZ.  
DR InterPro; IPR003165; Pw1.  
DR Pfam; PF02170; PAZ; 1.  
DR Pfam; PF02171; Pw1; 1.  
DR PROSITE; PS50821; PAZ; 1.  
DR PROSITE; PS50822; Pw1; 1.  
DR NON TER 808  
SQ SEQUENCE 808 AA; 91366 MW; 19528F4B9D10474C CRC64;

Query Match 44.2%; Score 1811.5; DB 5; Length 808;  
Best Local Similarity 49.4%; Pred. No. 8.9e-136;  
Matches 354; Conservative 125; Mismatches 228; Indels 9; Gaps 6;

QY 14 VKESKTSSGSGIVLSTHFRITSRPOVALQYHIDVNPMLFARLRSAALFQHEDLIGK 73  
DB 98 VKQALAGK--IALIANGFKTKTQYRVDFEPELINPRAFLIKGSHALGK 154  
QY 74 CHAFDGTLLPRLQKQYEVFSKTRNGEDVRITITLNELPPTSPTCLQFYNIIFRL 133  
DB 155 GLTLDMDTMYSLFKLAEKVTNLSARXKDSNVNICTHVAATINPMAPNTLHLNVLPFRC 214  
QY 134 LKIMTLQOIGRNYNPNPDIPIBSRLVYMPGFTTSIIQVENSIMLCQDVSHKYLRSSTV 193  
DB 215 LKIMIMEQGRNYDPTAIDIKQGLQMPGFTSIIQVYDVLSDISHKVLRQTV 274  
QY 194 LDFMENFHOEHEHFOEVSKEILGLVVLTKYNNKYRVDDIDMDQNKSTFKKADGSE 253  
DB 275 WEVWMDLFRKA-RGRFKEIETKMLGQIVLTKYNNKYRVDDIDFDTPADTFETRS-G-P 332  
QY 254 VSPLEYKQYKQNETLTKQPVYVQPKRRGPGTLLGPAMLLPELCYLTGLTDKMRND 313  
DB 333 VSYDYEFKSYERIVHDVQNPWLVSPPKREKGV--GPAYLPELCLFTGLSDMRAD 390  
QY 314 FVVMKDLAVHRTLPQROREVGRLIDYIHKNVQRELQWGLSFDNLSFSGRILOT 373

DB 391 FVVMKDLAVHRTVPQDRCTLSGFIKKLSNEEVKTYLDSWGMFDEKQYKLTGRVLP 450  
QY 374 EKHQGGKTEPDYNFOPADWSKETRGAPLISVYPDLNMLLYTRNYPEANSLIONLEKVT 433  
DB 451 EKLQGRKQFESYNNDMSDRTGNALTDKILNNWKIFYRRADNRGQDFIKSLVRA 510  
QY 434 PAMGQKRAIMIEV-DRTFAIYLVQOKYLTADTOIVGLSSNRKDKYAIKKYCTD 492  
DB 511 NPMGNVAGPEIIVLPDRITETYSLOAQIADTOIVLVLPTRKDRYAIKTCVYT 570  
QY 493 CPTSGQCVARTLGKQQTVAIAKRLALONCKRGELMRVDIPLKMTIVGIDCYHMT 552  
DB 571 HPCBQIVVSRITLSKQQLMSVATKIMQNNCKRGGLMRVEIILSNMIGIDSYHSL 630  
QY 553 AGRSIAGFVASINEGTRMFSCRIPQDRGOELVDGKVCLOALBANS CNEYMPRII 612  
DB 631 TKGRSVLGFVASNNKQTSFSSCAFOHAGEFGANLSTLMNNALKRYQINEKPEKII 690  
QY 613 VYRDGVDGQKTLVNEVDFLDCL-KSIGYNPRLTVIVKRVNTREFFAOGGRLQ 671  
DB 691 IFRDVGDSQYNLVVDYELKQIKOTLDKVPQGTVHKLAVVVKRIINRFFAMRGGLS 750  
QY 672 NPLPQTVIDVETREPTVDFIVSQAVSSGVSTHINVIYDNGSLKPDHQRILTY 727  
DB 751 NPPPGTVIDVETREPTVDFIVSQAVSSGVSTHINVIYDNGSLKPDHQRILTY 806

Search completed: May 5, 2004, 15:14:13  
Job time : 55 secs